DOCUMENT RESUME

ED 355 245 TM 019 581

TITLE	Performance Assessment & KERA. Proceedings of a Conference of the Kentucky Educational Research Association (Lexington, Kentucky, April 1991).
INSTITUTION	Kentucky Univ., Lexington. Center for Professional Development.
SPONS AGENCY	Eastern Kentucky Univ., Richmond.; Louisville Univ., Ky.; Western Kentucky Univ., Bowling Green.
PUB DATE	Apr 91
NOTE	61p.
PUB TYPE	Collected Works - Conference Proceedings (021)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Conference Papers; *Educational Assessment; Educational Research; Elementary Secondary Education; *Evaluation Methods; Humanities; Item Sampling; Language Arts; Mathematics Tests; Middle Schools; Program Implementation; Reading Tests; Second Language Instruction; Social Studies; *State Programs; *Student Evaluation; Teaching Methods; *Test Use
IDENTIFIERS	Alternative Assessment; Kentucky; *Kentucky Educational Research Association; *Performance Based Evaluation; Reform Efforts

ABSTRACT

In April 1991, members of the Kentucky Educational Research Association (KERA), held a conference to discuss performance assessment and its use in Kentucky schools. The following papers from the conference are included: (1) "Our Conference and the Performance Assessment Committee" (S. Kifer); (2) "Performance Assessment: A National Perspective--Cautions and Concerns" (G. W. Phillips); (3) "Elementary Education" (C. Bridge and A. Chapman); (4) "Elementary Education" (A. Zollman and J. Dickey); (5) "Middle School" (B. Burruss and R. Kennedy); (6) "Reading, Writing and Language Arts" (P. Winograd and J. Craig); (7) "The Practical Arts" (S. Miller); (8) "Arts and Humanities" (R. Schmelzer and G. Szekely); (9) "Reform Law" (V. D. Nordin and J. M. Petrosko); (10) "Implementation and Management" (K. Simpson, S. Trimble, K. Draut, and T. Mowery); (11) "Item Sampling" (G. Cunningham and S. Kifer); (12) "Performance Assessment for Special Populations" (D. B. Slaton and J. McLaughlin); (13) "Student Services" (P. de Mesquita and M. Billings); (14) "Mathematics" (W. S. Bush and S. Vice); (15) "Foreign Languages" (P. L. Nacke and L. K. Worley); (16) "Instruction and Assessment" (T. Guskey); (17) "Social Studies" (L. Levstik and S. Brennan); and (18) "Summary" (E. Mason). (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

Performance Assessment

&

KERA

ED355245

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
EMANUEL J. MASON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Sponsored By

Eastern Kentucky University

Western Kentucky University

University of Louisville

The University of Kentucky

Conference Proceedings

April 1991

TM019581

BEST COPY AVAILABLE

Performance Assessment

&

KERA

Conference Proceedings

April 1991

Acknowledgements

A conference so large involves many persons. We would like to thank the Program Committee for helping with the planning:

Linda Worley, Ron Atwood, Alan Zollman, Bill Bush, Frank Bickel, Wayne Harvey and Robb Fons from the University of Kentucky, Jim Craig and Rob Kennedy from Western Kentucky University, Joe Petrosko from the University of Louisville and Ann Chapman from Eastern Kentucky University.

Those who conducted sessions included:

Ron Atwood, Jim Craig, Mike Howard, Bill Bush, Sheila Vice, Linda Levstik, Sharon Brennan, Joanna Dickey, Alan Zollman, Connie Bride, Ann Chapman, Linda Worley, Phil Nacke, Betty Burris, Rob Kennedy, Peter Winograd, Sandy Miller, George Szekely, Ron Schmelzer, Virginia Davis Nordin, Joe Petrosko, Eddy Van Meter, Linwood Doake, Tom Mowery, Kawanna Simpson, Scott Trimble, Ken Draut, Ed Blackhurst, Leroy Metze, George Cunningham, Debra Bott Slaton, Jay McLaughlin, Lesa Billings, Paul DeMesquita and Tom Guskey.

The afternoon panel session included Connie Bridge of the University of Kentucky, Charles Campbell, President of KASS, Ben Oldham of the Fayette County schools and Gary Phillips, our excellent keynote speaker from the National Center for Education Statistics, a branch of the Office of Educational Research and Improvement.

A special thank you goes to those who welcomed the group: Robert Hemenway, Chancellor Lexington Campus, J. John Harris III, Dean of the University of Kentucky College of Education, John Brock, Superintendent of Public Instruction and Thomas Boysen the Commonwealth's new Commissioner of Education.

The Center for Professional Development in the College of Education at the University of Kentucky organized the conference. Our thanks to Jo Omvig, Robb Fons and Wayne Harvey.

And finally our thanks to the participants for making the conference a good one.

Emanuel Mason - Skip Kifer

April 2, 1991

Contributors

Gary W. Phillips	*National Center for Education Statistics*
Connie Bridge	*University of Kentucky*
Ann Chapman	*Eastern Kentucky University*
Alan Zollman	*University of Kentucky*
JoAnna Dickey	*Eastern Kentucky University*
Bette Burruss	*Western Kentucky University*
Rob Kennedy	*Western Kentucky University*
Peter Winograd	*University of Kentucky*
Jim Craig	*Western Kentucky University*
Sandra Miller	*University of Kentucky*
Ronald Schmelzer	*University Of Louisville*
George Szekely	*University of Kentucky*
Virginia Davis Nordin	*University of Kentucky*
Joseph M. Petrosko	*University of Louisville*
Kawanna Simpson	*University of Kentucky*
Scott Trimble	*Kentucky State Department of Education*
Ken Draut	*Kentucky State Department of Education*
Tom Mowery	*Kentucky State Department of Education*
George Cunningham	*Unversity of Louisville*
Skip Kifer	*University of Kentucky*
Deborah Bott Slaton	*University of Kentucky*
Jay McLaughlin	*University of Louisville*
Paul de Mesquita	*University of Kentucky*
Malessa Billings	*Kentucky State Department of Education*
William S. Bush	*University of Kentucky*
Sheila Vice	*Kentucky State Department of Education*
Phil L. Nacke	*University of Kentucky*
Linda Kraus Worley	*Univeristy of Kentucky*
Thomas Guskey	*University of Kentucky*
Sharon Brennan	*University of Kentucky*
Linda Levstik	*University of Kentucky*
Emanuel Mason	*University of Kentucky*

Table of Contents

Our Conference and the Performance Assessment Committee
Skip Kifer 1

Performance Assessment: A National Perspective - Cautions and Concerns
Gary W. Phillips 3

Elementary Education
Connie Bridge *Ann Chapman* 9

Elementary Education
Alan Zollman *JoAnna Dickey* 13

Middle School
Bette Burruss *Rob Kennedy* 16

Reading, Writing and Language Arts
Peter Winograd *Jim Craig* 18

The Practical Arts
Sandra Miller 21

Arts and Humanities
Ronald Schmelzer *George Szekely* 22

Reform Law
Virginia Davis Nordin *Joseph M. Petrosko* 24

Implementation and Management
Kawanna Simpson *Scott Trimble* *Ken Draut* *Tom Mowery* 26

Item Sampling
George Cunningham *Skip Kifer* 30

Performance Assessment for Special Populations
Deborah Bott Slaton *Jay McLaughlin* 32

Student Services
Paul de Mesquita *Malessa Billings* 34

Mathematics
William S. Bush *Sheila Vice* 37

Foreign Language
Phil L. Nacke *Linda Kraus Worley* 39

Instruction and Assessment
Thomas Guskey 41

Social Studies
Sharon Brennan *Linda Levstik* 50

Summary
Emanuel Mason 52

List of Participants

Our Conference and the Performance Assessment Committee

Skip Kifer

As I think most are aware there is a rather short time frame for implementing performance assessments in Kentucky schools. The legislation calls for an interim state-wide assessment in 1991- 1992 and a "full-blown" assessment in 1993 - 1994. In addition, there are requirements for implementing simultaneously a program of continuous assessment in each of the Commonwealth's schools.

The legislation also sets a procedure for designing and implementing the assessment program. Among the procedures are: a committee of assessment experts is to write a Request for Proposals (RFP) that contains a framework for the assessment; a contract is to be awarded to a successful bidder who will carry out the assessments, and; the successful bidder will work with schools to produce the results of the assessment.

I am a member of the Committee that is to write the RFP and that is what I would like to talk about briefly . I hasten to say that I can speak about the committee and its work but I cannot speak for the committee. Other members of that committee include: Pat Forgione who presides over and conducts in Connecticut what is considered to be an exemplary state assessment program; Jason Millman, from Cornell university, whose expertise is in the technical arena and experience includes having served on the technical committee for the California state assessment; Doris Redfield who is an evaluator from the Center for the Study of Evaluation at UCLA and who has worked closely in the recent past with Kentucky schools; Grant Wiggins who is the well-known proponent of performance based assessment.

Members of the committee are familiar with major assessment programs in the United States (California, Connecticut, New York, Vermont) as well as those which are well-known internationally (e.g., Great Britain and Australia). They are, I believe, sensitive to both the potential for these kinds of assessments and the problems they produce.

The Committee was appointed in late November and has met twice to discuss the framework for the assessment and the structure of the RFP. There are plans to meet again later this month.

The Committee has spent time understanding the nature of the legislation and the requirements of the Kentucky assessments. It has benefited from discussions with those who framed the law and those who are working on other aspects of the reform which have implications for performance testing. For example, we talked with Roger Pankratz, the leader of the task forces working on framing goals and defining measurements for the assessment, on both occasions that we met. In addition, Roger has made sure that there has been a continuous flow of information from his groups to our committee.

Because the Kentucky legislation has components similar to those in other states, neither the committee nor the Commonwealth will be starting at point zero in terms of our assessment. Existing practices in other states and countries should provide a basis for framing the state-wide performance assessment.

In addition to obvious similarities between our legislation and other already existing programs, the Kentucky assessment has features which make it unique. Of particular note (and something you must certainly be aware of) is tieing the state-wide assessment to school accountability. That feature of the law makes this assessment different from other existing assessment programs. It also gives the Commonwealth an opportunity to lead other states as they introduce performance assessment programs.

Issues of performance assessment in the Commonwealth are complex and exciting. Because the RFP has not yet been written and, obviously, a contractor has not been chosen, the precise form of the assess-

ment is not known. The structure of the assessment program, the content to be assessed and the methods of assessment are open in one sense. In another sense, there are some obvious precedents in the work of our task forces and existing practices in other states and countries.

So, although there is ambiguity surrounding Performance Assessment in the Commonwealth, there are things we can do to prepare for program when it is implemented.

We hope this conference will help you with that extraordinarily important work. Our morning sessions are designed to provide information about how performance assessments have been carried out in other arenas for various content areas and levels of schooling. We hope they will help you build a larger and more fully formed apperceptive mass about performance assessment. That apperceptive mass, a repetoire of information and insight will, we assume, help you respond to the assessment when it is precisely defined and implemented. The more we know; the better we will do.

I said earlier that the committee has had an opportunity to converse with those who have framed the law and those who at the state level are working on matters that are related to the assessment. It has not conversed with those in the schools who will implement this program. Your second session of the day focuses on your insights and views about barriers to implementation. Those insights and views, along with other parts of today's program, will be summarized in a volume of proceedings. I will take that volume, containing those insights and views, to the committee. This your second task will, we hope, make the assessment committee's task easier. The committee welcomes your thoughts.

Performance Assessment: A National Perspective - Cautions and Concerns

Gary W. Phillips

Thank you very much for inviting me to speak at this exciting conference. There is much happening in educational reform all around the country and Kentucky is certainly the eye of the storm. The whole country is watching to see how your reforms turn out. They will certainly serve as a model for others to follow.

I would like to talk about two of my favorite topics. The first topic will be NAEP and how your activities might piggy- back on what we are doing at the national level. The second topic will be what I think are some lessons I have learned in my work on performance assessments at the National level that you should keep in mind as you develop your performance based testing program. I want to warn you that I'm one of those evil psychometricians that your testing program will eventually have to deal with.

First, let's talk about NAEP. As you may know the NAEP has been around for over 20 years monitoring the educational achievement of our nation's school children. We recently published a 20 year summary of findings and the picture looks pretty bleak. The overall finding is that in the past 20 years there has been little or no improvement in any subject area, or at any grade. This is in spite of the large number of educational reforms and large increases in per pupil expenditures.

In 1990, NAEP conducted its first Trial State Assessment in 37 States, two territorities and Washington, DC. The results will be released on June 6, 1991 in a national press conference and 40 simultaneous State press conferences. Kentucky was one of those participating States. The results will cover mathematics at the 8th grade level for public schools only. In 1992 the Trial State Assessment will be extended to also include 4th grade reading and mathematics. Kentucky has also agreed to participate in 1992. Our National Assessment Governing Board is also setting educational standards for the first time. When they are finished we will be able to provide State-by-State comparisons on how many students are achieving at a basic, proficient and advanced level. This will be the first time NAEP has gone beyond measuring what students know and can do to how many children have learned what they should know and be able to do.

The Kentucky Education Reform Act requires you to pay attention to what is happening in NAEP. The Kentucky legislation requires the State to create an interim NAEP-like testing program beginning in the 1991-1992 school year that is representative of each school. Let me discuss several ways this might be done. At the present time NAEP cannot be administered in such a way that student, school or district results are permissible. This is due to the prohibition in Federal legislation. First let's talk about "What does NAEP like mean?" I'm not sure what you mean by the term but here is what I mean by the term. 1)NAEP objectives and item specifications are used in developing your test. 2)Matrix-sampling is used such that each student in the class gets a portion of the test but no student takes all of the test. 3)Demographic and Contextual information is collected from students, teachers and principles to help explain student academic performance. 4)The results are expressed on the NAEP scales. 5)The test yields reliable information for schools and districts but does not provide reliable information for individual students.

In spite of the legislation there are at least two ways that representative school type data can be obtained and still get national NAEP comparisons. The first would be to develop NAEP- like tests based on a combination of released NAEP items supplemented by writing new NAEP-like items developed according to the publicly available NAEP objectives booklet. As an example 3/7 of the 1990 Trial State Assessment items have already been released to the

public, and would be available for such a strategy. The link to the NAEP national scale would be possible through equating studies. The equating studies would be conducted (at the State's expense) in conjunction with the NAEP assessment. The next opportunity for this would be in the winter of 1992. A second approach would be to develop your own State testing program but have a portion of it replicate the NAEP. Again through equating studies at least that portion of your State testing program could be equated to NAEP. This would yield NAEP equivalent scores for a portion of your assessment while preserving the individual identity of your overall testing effort.

Several commercial testing companies are looking into the profitability of both approaches and the University of Tennessee is seeking initial funding from several foundations to develop such procedures. These are at least two approaches to fulfilling your legislative mandate. There may be other creative approaches. One thing to keep your eye on is the NAEP re-authorization that will likely occur next year. One of the issues Congress will deal with is that several States have passed legislation requiring their State testing to use NAEP, or NAEP-like tests.

Let me now turn to the topic of authentic assessment.

Let me begin by again confessing that I am a psychometrician. I also think I'm a good psychometrician. Psychometricians these days are like proctologists. No matter how good you are, you still get no respect.

While standing in the hallway the other day I heard two experts on authentic assessment tell a joke. The joke was "what do you say about a psychometrician who is up to his neck in cement?" "More cement!"

Since the beginning of the performance assessment movement, several years ago, psychometricians have gone into the closet. They have been branded as establishment conservatives who are more concerned with how the numbers come out than how the students come out. All around the country and even the world they are treated as the educational equivalent of lawyers. First kill the lawyers then we will have a better society.

Let me begin my remarks by letting you know something about my experience with performance assessments. My more recent experience has been with the National Assessment of Educational Progress (NAEP).

For 20 years NAEP has assessed writing using direct writing methods. We were certainly one of the first testing programs to assess writing in an authentic way. Our current writing test requires students to write for 25 to 50 minutes and we score the test using analytic, primary trait and holistic scoring methods.

Our Adult Literacy Assessment in 1985-86 (to be repeated again in 1991) was also an entirely authentic test. There were no multiple-choice items. The test involved one-on-one interviews with the use of actual checkbooks, newspapers, bus schedules, maps, etc. Our current Math assessment on which 40 States and territorities are being compared involves the use of rulers, protractors, and 30% of the items require scientific calculators. Thirty percent of the items are open-ended and the entire assessment incorporated the NCTM views on how math should be assessed.

Our upcoming reading assessment involves reading passages taken unedited from the literature and 40% of the students time is spent writing in response to open-ended questions.

In past assessments NAEP used hands on tasks in science, assessed children's drawing skills, assessed students' ability to sing and play musical instruments, assessed dynamics of group decision-making and assessed students' ability to write computer programs.

I mention these facts about NAEP because I hope to convince you that I'm a believer in the benefits and promises of performance assessments. Performance assessment strategies may turn out to be the greatest advance in testing in recent decades. Several States have taken the plunge (California, Connecticut, Vermont, Maryland and Kentucky) and represent the leaders in the field. Performance assessments hold the greatest promise of assessing effectiveness of instruction, and keeping testing and curriculum aligned. However, I do think the benefits of performance assessments are more likely to be realized in classroom, district and possibly State level testing programs. As the scale of the assessment gets larger (national and international) it becomes more difficult (but not impossible) to realize these benefits. Although performance assessments clearly represent a major advance in the testing business, there are many exaggerated claims about their benefits, and many misconceptions that someone needs to talk about. When asked to speak at this conference I decided the most useful service

I could provide would be to recommend some cautions based on our experience at the national level which will require your attention as you proceed with implementing a statewide performance assessment. I hope this will help to round out the advice you have been getting on the subject.

The first caution is that advocates of performance assessments argue that multiple-choice items cannot measure problem solving, reasoning and higher thinking skills. They cite evidence from NAEP and various international assessments that demonstrate that students are not learning these higher level skills. They argue that the reason students aren't learning these skills is because we aren't testing for them. But wait a minute folks. All of the evidence that students lack these skills is based on assessments that use primarily multiple-choice items. NAEP used multiple- choice items to show that in the past 20 years we have improved in basic skills and dropped in higher order thinking. International studies that have come to the same conclusion use almost exclusively multiple-choice items. It is true that performance assessments possibly provide better measures of higher-order thinking skills, but don't forget that we also want to know how many students only have basic skills. The jury is still out on how good performance assessments are at measuring basic skills.

A second caution is that performance assessments are said to somehow lead to more equitable testing because minority students have a better chance to show what they have learned in a real world context which respects divergent ways of thinking, and creative solutions. Proponents of performance assessment envision a time when the skills of minorities will finally be tapped and the majority/minority gap in educational performance will finally be closed. This may be true, I hope it is, but what does the evidence say? Based on preliminary data we have noticed the opposite in the National Assessment. The gap between White and Black students, and White and Hispanic students gets larger as the item format gets more authentic. In fact in some cases high proportions of the minority students get zero on all of the more difficult authentic assessment items.

A third caution is that many people argue that performance assessments allow you to measure richer, fuller, more rewarding content. It is not always clear what this means but I think the argument confuses two desirable but competing content considerations - depth versus breadth. Performance assessment allows you to measure a few things in great depth but more traditional methods will provide a greater scope of coverage. Traditional methods of criterion- referenced testing achieve the greater breadth of coverage by using domain sampling methodology. In this method, the test is only a random sample of items from a larger domain (such as Geometry). By randomly sampling items, the test scores can be generalized to represent content not included on the test. In performance assessments the concept of domain sampling is often abandoned. Here the focus is on the few good items that capture the big ideas, the thought processes, the actual skills you want the students to learn. In performance assessments, inferences beyond the items are tenuous. If the test includes the entire universe of things you want the student to know then you are in luck. However, in most cases it is not practical to test students on everything you want them to know. This is why domain sampling theory was invented in the first place.

The third and I think the most important caution about performance assessments is that they are claimed to be more valid then multiple-choice tests. Supporters of this view cite the fact that authentic assessments sample the actual behaviors and skills we want students to learn. This is given as prima facia evidence of validity. But this is what we psychometricians call face validity, or more skeptically, faith validity. You simply cannot tell how valid a test is by looking at it, or liking it. Nor can you make the test more valid by rhetorical exercises of creating new forms of validity (such as ecological validity and boy-this-is-a-great-test validity).

The concept of validity is the most important concept in the testing business. People who claim they are testing professionals must understand this concept and must subscribe to the professional standards that define the concept and prescribe how it should be measured. Because the concept is so important I would like to spend much of my time talking about the concept and, if nothing else, I hope that I leave you with just one thing -- a better understanding and appreciation of the importance of validity and an appreciation of how performance assessments must take the concept more seriously.

Validity refers to the extent to which inferences from a test are appropriate. It is not the test that is valid, it is the inference. The joint technical Standards for Educational and Psychological Testing (by

AERA, APA, and NCME) define validity as the "appropriateness, meaningfulness, and usefulness of the specific inferences made from the scores. Test validation is the process of accumulating evidence to support such inferences ... validity is a unitary concept (and) ... always refers to the degree to which that evidence supports the inferences that we made from the scores. The inferences regarding specific uses of a test are validated, not the test itself" (p.9). So in what sense can performance assessments be more valid than multiple-choice tests? They cannot be more valid in the sense of greater face validity because face validity is not a form of validity. Face validity is a concept outside the joint technical standards and refers to the political acceptance or the appeal of the test based on its superficial appearance. It is abundantly clear that performance assessments have greater face validity then multiple- choice tests. However, that doesn't mean they are more valid in the sense of providing better inferences.

One way in which performance assessments might be argued to be more valid is that they are more representative of the defined universe or domain of content we wish to test. That is they have greater content validity because they better represent the actual skills and behaviors we want to learn. They are samples of the most important parts of the curriculum on which there is expert consensus, and what we want teachers to emphasize.

Although performance assessments hold the promise of being more content valid, the evidence for such validity is often asserted but seldom demonstrated. To be content valid means that there is a universe of content to which we wish to make inferences from the student's test scores. For example, we may wish to make an inference about a student's narrative writing skills from his score on a writing prompt, or we may wish to make an inference about a student's understanding of the scientific method from a science project. But how can these inferences be valid if the writing skills of hundreds of thousands of students are based on only one or two writing prompts (a one or two item test - the norm for virtually all performance testing programs) and how can you compare hundreds of thousands students on how much they understand the scientific method if each is doing a different science project? The issue is not that hundreds of thousands of science projects can be evaluated by trained evaluators with rigid rating criteria. The issue is whether the inferences are equally valid for each of the hundreds of thousands of students. If they are not, then the testing program is unfair to students for whom it is less valid, especially if high stakes decisions, such as promotion or graduation, are a consequence of the scores.

The problem of equal validity is exacerbated when alternate forms of the testing instrument must be developed in subsequent years. If a reading assessment uses a well known short story such as "A Tell-Tale Heart" by Edgar Allan Poe, then teachers will start teaching the short story. Since we want to teach lots of short stories (not just the one on the test) we must use a different short story on the test the following year. But it is hard to find another short story which has all the properties we liked which led us to select "A Tell-Tale Heart" in the first place. And what about the next year, and the next. If we choose short stories that are different form "A Tell-Tale Heart" it is unfair to measure subsequent students by a different yardstick. This is the psychometric problem of equating in authentic assessments. Anyone can conduct a good assessment once, the real trick is to do it twice.

In the final analysis the greatest obstacle to the validity of authentic testing is unreliability. Unreliability is the Achilles' heel of performance assessments. Eventually the problem must be dealt with in a responsible way. We are wrestling with it at the national and international level, and Kentucky will have to deal with it at the State level. Let me take a minute to explain what I mean. There is a mathematical relationship between the reliability and the validity of a test. It is impossible for the validity to exceed the square root of the reliability. This means that an unreliable test cannot possibly be valid. There are two reasons why performance assessments are generally less reliable than multiple-choice type tests.

The first is that authentic tests typically employ fewer items. Because of the cost of scoring and the time burden on students, most authentic assessments employ so few items that their reliability is suspect. Some State assessments of writing use only one item. Under these circumstances the reliability of the test cannot even be estimated (this requires at least two items).

The second major reason why performance assessments tend to be less reliable is because they cannot be administered under standard conditions and they cannot be objectively scored by a computer. The

administration and scoring of the assessment permits many sources of uncontrolled variation. The test is administered under more realistic real-life, uncontrived conditions. The students' responses are scored by raters who vary in their grading leniency. To better understand this it is helpful to imagine an analogy where performance assessments are used routinely. We often hear that performance assessments are similar to what a football coach does to evaluate a player. But in a large scale assessment the analogy is more complicated. Imagine thousands of football coaches evaluating hundreds of thousands of players, using thousands of different plays which are chosen by the players, not the coach. This is identical to a portfolio assessment or an assessment of science projects. With so many uncontrolled variables it becomes very difficult to make objective (i.e., equally reliable and valid) decisions about the students' performance.

I would like to end my talk with some free advice. My advice is basically that Kentucky needs to spend a lot of time understanding the practical and technical measurement problems associated with starting a new, innovative testing program. The program will affect the lives of countless thousands of students, parents and teachers. Furthermore, you are doing it in a fish bowl. The whole country is watching to see if you can pull it off. You've made the decision that you are going to the moon. Now everyone is watching to see if you get there. I have five specific recommendations.

Recommendations

- Think through measurement issues. Kentucky has spent a lot of time and resources focusing on curriculum and instruction issues. Spend some time on measurement issues. After all testing is a measurement activity. A large scale testing program must address hundreds of technical and theoretical measurement issues that go beyond the curriculum. You must build a testing program that responsibly addresses these problems. Some of the technical issues include test bias, scaling, cost, standardized administration, reliability, validity, equating, and scoring. The theoretical issues center around the purpose of testing. What are you trying to do? Are you trying to monitor performance, trying to improve performance, hold teachers and their schools and districts accountable, make decisions about resource allocations, evaluate programs, promote or graduate students, give diagnostic feedback to teachers, evaluate instruction, evaluate curriculum? One test cannot do all of these things. Decide up front what you are trying to do.

- Read carefully and conform to all of the joint technical standards written by APA, AERA and NCME. These represent the best thinking by the best people in the field about your professional testing responsibilities. Your testing program must meet and hopefully exceed the standards of the profession. When your actions start to actually impact teachers and students you will need the professional standards to corroborate your procedures. If there are legal challenges you will have support for your actions. If you do not follow the joint technical standards then I suggest you keep your resume current.

- Now my most heretical recommendation. Don't completely abandon the multiple-choice format. Try out the new and more innovative approaches to assessment, but build in an honest evaluation system that will help you make appraisal of your success, and give yourself time to make mid-course adjustments. Take a risk, but remember the fundamental economic principal in any investment -- the greater the risk the greater the need for diversification. Try out performance assessments. But don't forget the lowly multiple-choice item. It has been around for over a hundred years performing the thankless task of providing reliable and valid data in a cost effective manner. I suggest you include it in your assessment portfolio.

- Get the very best people to work on the project. Not only in its initial stages, but even more importantly, for the long

haul. Long after the visionaries have gone on to other callings, long after the consultants and contractors have set the system in place, there will be a group of nameless, faceless, but loyal civil servants who like the multiple- choice item, will be called upon to make the system work.

- Finally, don't be deterred by people like me who say it cannot be done. It can be done. But be prepared for a difficult task. The last time we said we were going to the moon we made it. Even though it was a complicated technological task the basic science was in place to get us there. What was needed was the will and resources. We are in an identical situation with performance assessment. The will is there, the science is there. What we need now are the effort and the resources. What we have learned at the national level is that performance assessments involve a lot of people, take a long time and cost a lot of money. If Kentucky is willing to provide the effort and resources then I wish you well on your way to the moon.

Elementary Education

Connie Bridge

Ann Chapman

We began the session by discussing the following definitions of performance assessment and the criteria for high quality performance tasks currently being used by the Council on School Performance Standards to guide the task forces in the development of performance tasks for Kentucky schools:

Performance-Based Assessment is defined as giving a group of students, or an individual student, a task that may take from half an hour to several days to complete or solve. The object of the assessment is to look at how students are working as well as at the completed tasks or products. An observer or interviewer may stay with the group or make periodic visits. Activities may be videotaped, tape recorded, or recorded in writing by an adult or students. The task might be from any content area and might involve any subject (Lee and Lee, 1990).

Roger Pankratz (1990), Executive Director of the council of School Performance Standards, states that "performance assessment measures a student's ability to perform a task like writing a business letter, balancing a checkbook, performing a science experiment, using a word processor to edit a manuscript or repairing an appliance. Performance assessment usually involves measuring a student's ability to use knowledge and skills to produce an outcome or product . . . [and thus] looks beyond the knowledge and skills we usually measure with paper and pencil tests to what students can do with the knowledge and skills they have acquired."

Drawing upon the work of Grant Wiggins, Pankratz has suggested that the following criteria be used to evaluate the performance tasks currently being developed in Kentucky:

- Essential vs. Tangential - The task fits into the core of the curriculum and represents the "big ideas" of the discipline.
- Authentic vs. Contrived - The task assesses skills, knowledge and processes that are most appropriate to, valued in, and commonly used in the discipline. The task is worth learning and teaching and simulates the work of adults in a given field of study.
- Rich vs. Superficial - The task leads to other problems and raises other questions; it is "pregnant with possibilities."
- Engaging vs. Uninteresting - The task is thought-provoking, challenging and fosters persistence.
- Active vs. Passive - The task encourages student activity and interaction with other students and involves the student in constructing meaning and deepening understanding.
- Feasible vs. Infeasible - The task can be performed with available resources and completed within reasonable time constraints.

Assessment tasks which exemplify these criteria will drive instruction in positive ways. They will encourage teachers to focus on integrated application of important learnings rather than on memorization of isolated skills. They will encourage students to think through the processes they are using and to explain why and how they achieved results.

Fortunately, our best teaching practices are already preparing students for this type of assessment. Examples of these practices include integrated, holistic reading and language arts instruction; writing across the curriculum; hands-on, manipulative experiences in mathematics and science; and cooperative learning.

We discussed Goals 3 and 4 of the Six Learning Goals (Council on School Performance Standards, 1990). Goal 3 stresses Self-Sufficiency and Goal 4 stresses Group Membership Skills. Most par-

ticipants agreed that the best way to assess these two goals is within the context of the tasks designed to assess Goal 1 Basic Skills and Goal 2 Core Concepts. Many of the desired outcomes related to self-sufficiency and group membership (e.g., learns on his/her own and recognizes the viewpoints of others) could be observed while students were performing either group or individual tasks in the basic content areas. However, some members of the group expressed concerns about whether group dynamics could negatively affect the evaluation of individual performance, and might interfere with the evaluation of the student's understanding in a given content area.

Teachers will need to learn how to facilitate cooperative learning groups if students are to be prepared for working effectively in group situations. This was viewed as another example of one of the ways in which the new performance assessment will drive instruction in positive ways and provide affirmation for the teachers who are already engaging in innovative teaching techniques.

All participants agreed that the criteria for a good performance should be agreed upon in advance and shared with the students ahead of time. It is important that both teachers and students be made aware of what constitutes an expert performance on a given task so that the teachers will be able to prepare students for the task and students will be able to engage in self-evaluation.

Participants expressed strong concerns about the validity and reliability of performance assessment. Although the individual tasks have high face validity, time constraints will make it impossible to sample broadly from the total discipline because each task will take a long time to administer. For example, if students are tested only on one or two tasks related to history, will the results of that assessment provide an accurate reflection of the students' historical understanding? If there are strong within-student differences among tasks in a content area, the results may present a misleading picture. Will the trained assessors be able to make reliable judgments after observing students for only a short time in a limited number of situations?

While we agreed that authenticity certainly contributes to the face validity of a task, there is a danger that tasks that are truly authentic in one geographic area of the state might not be authentic in another. For example, a math question regarding toll booths might not provide a familiar context for students in Eastern Kentucky who have not traveled in the western part of the state.

Furthermore, there were questions regarding the matrix sampling procedures to be used in selecting the students from a given school who will be asked to participate in a particular performance assessment task. Will enough students be assessed to give a true picture of the performance of the total school in a given discipline?

Several questions were raised about the Interim Test to be administered in 1991. If the Interim Test is to resemble the NAEP (National Assessment of Educational Progress) tests rather than the Kentucky Performance Assessment Tasks, how will the state be able to set the threshold level on the Interim Test and then judge school performance by a retest on the Kentucky Assessment? Will some schools purposely score low on the Interim Test so that they can show improvement on the next testing? As soon as the state has made decisions regarding the Interim Tests, educators need to be informed immediately to enable them to prepare for the assessments. Presently, anxiety levels are high.

Overall, participants felt that the positive effects of performance assessment will likely outweigh the negative in that it will drive instruction in positive ways and keep the focus on learning processes and authentic products rather than on isolated skills and will affirm the efforts of our most innovative teachers who are already stressing cooperative learning and developmentally appropriate curriculum. However, participants expressed strong needs for more information about the nature of the performance assessment to be used in Kentucky and for staff development to help teachers learn how to prepare their students for this new type of assessment.

Afternoon Session

A. Do the activities defined, illustrated and discussed in the morning session meet the requirements of the law? If not, what must be done to create a fit?

Due to the fact that the afternoon participants had attended different session during the morning, there was little common ground in this session for answering this item. Nevertheless, the discussion revealed that neither participants nor leaders could list with certainty the requirements of the law in the area of performance assessment.

Lack of clarity and specification can be positive as well as negative. It is often part of any new endeavor and can be an opportunity for individualization and creativity. However, our discussion indicated that in this instance the lack of clarity also involves a lack of direction and basic knowledge regarding performance assessment.

As difficult as it may be, those individuals directing performance assessment in Kentucky need to concretize guidelines regarding what performance assessment means in the Commonwealth. This statement is made with full recognition that a final determination may only be possible with the evolution of the product. Nevertheless, as soon as possible, everyone involved with public education in Kentucky needs to be provided with basic knowledge, guidelines, definitions and samples regarding the range of what performance assessment is and what it may be like in the Commonwealth.

B. What are the structural barriers, if any, to using these activities in the schools?

Session participants listed the following potential barriers:

- Use of textbooks which do not facilitate the preparation of students for performance assessment.
- Inadequate preparation of older students to integrate material from different subject areas.
- Short length of time for implementation of the new approach to assessment.
- Task content
  - Who will select tasks? Will there be input from teachers?
  - Will there be an appeal process for districts/schools who disagree with the content, scoring, criteria, etc. of a task?
- Sampling
  - How will the matrix be developed and implemented?
  - What steps will be taken to insure that samples are not biased?

Participant discussion of these barriers revealed the high level of anxiety regarding performance assessment felt by both teachers and administrators. Participant statements indicated that much of this anxiety was due to a lack of knowledge regarding ways to begin to prepare for the new assessment. Thus, the importance of training, a topic which is discussed more thoroughly below, was clealry indicated by this group's discussion of barriers. Indeed, this group's discussion suggested that a lack of adequate training would be a major barrier to the implementation of performance assessment.

C. What are the Costs of Implementing these new Activities?

Our group did not respond to this question in terms of dollars and cents. However, they did indicate that costs will occur in several areas including test development, test administration, school operations and school functioning (hidden).

In regard to the tests themselves, the development, pretesting and revision of the assessment tasks will require extensive preparation and work. To carry out the state wide administration of the performance assessments, a large number of full-time personnel must be hired. These task administrators should be professionals hired at professional salaries. The administration of the tasks will also require a large travel and materials budget to enable the task administrators to do their job.

School related costs will include a budget line to support training (see below), and one for the purchase of materials, texts and books which will enable teachers to be effective as they teach for a new type of assessment. Hidden costs for schools may include the loss of some good teachers who are near retirement. Schools may also incur stress related costs such as an increase in sick days, low energy levels among faculty and administrators and an increase in the number of staff disputes.

D. Are There Areas in which Training Will Be Required?

The need for training was evidenced in almost every area discussed during the afternoon session. We cannot begin to implement the requirements of the law until we understand how performance as-

sessment will be defined in the Commonwealth. Both teachers and administrators must become familiar with the nature of performance assessment in general and performance assessment in Kentucky. (Training in the former can be done right now, training in the latter should commence as soon as possible.)

Teachers also need training which will help them understand what they can do in the classroom to prepare their students for performance assessment. Part of this training can focus on simply reminding teachers of techniques with which they are already familiar. Some of the suggestions which were made during our afternoon session were a) think of activities which motivate students and help them to learn, b) to talk to teachers who have hands-on activities in their classrooms, c) adapt performance based approaches to education, such as Montessori, d) use the principles of teaching for transfer, and, e) mentor and learn from new teaches with creative ideas.

A third group which will need training regarding performance based assessment is parents. If they are to support their children and the schools, they need to be aware of the manner in which their children will be evaluated and the skills which their children need to develop. This training should be introduced during the year preceding the first performance assessments.

It is important that school councils/administrators recognize and support the need for training. This will require the allocation, for both faculty and staff, of funds and time for inservice sessions and attendance at meetings and workshops. Each district/school must also determine the best way to educate parents, and insure that as many parents as possible are informed regarding performance assessment. Districts/schools should also be aware that training in the areas of communication skills and the management of stress may be very helpful during the implementation of KERA.

Finally, it is extremely important that state and regional education related organizations include sessions regarding performance assessment during their annual meetings. These meetings are affordable and may provide one of the most cost effective ways to provide training.

E. Session Summary

Participants in the afternoon session on elementary education clearly indicated the need for clarity regarding performance assessment in Kentucky. Barriers to the implementation of performance assessment include time and means to prepare students for a new type of assessment, task content, administration sampling procedures and generalized anxiety. Two keys to minimizing the barriers and successfully implementing performance assessment are training for faculty, administrators and parents, and adequate funding for all aspects of the program.

References

Council on School Performance Standards. (1989). Preparing Kentucky Youth for the Next Century: What Students Should Know and Be Able to Do and How Learning Should be Assessed.

Lee, M.M. & Lee, L.N. (August, 1990). Workshop conducted for Powell County school administrators, teachers, and staff.

Pankratz, R. (1990). What is performance assessment? Unpublished guidelines for the task forces of the Council on School Performance Standards.

Elementary Education

Alan Zollman

JoAnna Dickey

The Kentucky Education Reform Act of 1990 set the following goals for the Commonwealth's public schools:

1. Schools must expect a high level of achievement of all students.
2. Schools must develop their students' ability to:

   - use basic communication and mathematics skills for situations they will encounter throughout their lives;
   - apply principles from mathematics, sciences, arts, humanities, social studies and practical living studies to situations they will encounter throughout their lives;
   - become self-sufficient individuals;
   - become responsible members of a family, work group or community;
   - think and solve problems in school situations and in life;
   - connect and integrate experiences and new knowledge with what they have previously learned, and build on past learning experiences to acquire new information through various sources.

3. Schools are to improve their students' rate of school attendance.
4. Schools are to reduce their students' dropout and increase retention rates.
5. Schools are to reduce physical and mental health barriers to learning.
6. Schools are to be measured on the proportion of students who make a successful transition to work, post-secondary education and the military.

(KERA, 1990).

One aspect of KERA charges the State Board for Elementary and Secondary Education to develop a statewide assessment program, including performance-based student testing, to measure the success of each school according to the above goals (Legislative Research Committee, 1990). Students, beginning in the 1993-94 school year, will be tested in reading, mathematics, writing, science, and social studies in grades four, eight, and twelve. Local school boards will be required to publicize an annual report of performance of students. Thus, elementary schools will be expected to do performance-based assessments of their students.

What is performance assessment? What exactly do we mean?

Performance based assessment measures a student's ability to use knowledge and skills to perform a challenging activity or task. This implies that the assessment looks beyond mere knowledge and skills to actual application. Performance assessment requires a set of criteria for evaluating the quality of a performance (Wiggins, 1989: Pankratz, 1990). The attributes for the set of criteria include essential, authentic, rich, engaging, active, and feasible (Connecticut State Department of Education, 1990). Methods for the evaluations can include performance tests, observations, exhibitions, interviews, and portfolios (Mitchell, 1989).

By its name, assessing performance implies students performing, or "doing" something. What students "do" should be tasks that approach real-life situations. To assess these tasks seems to require some type of observations. But is it necessary to be observed? Are certain skills even observable? If the tasks are to be appropriate, the tasks could thus require a lot of time to complete. Will not this time, therefore, present more problems? Who will do the observations? How will they observe? What will be the criteria?

Along with observations, student portfolios also seem to be appropriate to assess performance. Here again, logistics will be a strong concern. How will

these portfolio records be compiled, produced, and stored? Also, what will be required of the content of the portfolios?

Performance-based assessment tends to be viewed as either promoting convergent thinking or divergent thinking, depending on the point of view of the educator. If, in order to assess performance, scoring rubrics are stated, how will creativity and resourcefulness be included? Upon whose criteria will the scoring rubrics be constructed?

What are the criteria for determining whether performance assessments are good or bad? When comparing performance assessments to more traditional paper and pencil measures are there issues of validity or efficiency?

For "authentic" performance-based assessment, logistics can be a major concern with issues of validity. Reliability will be difficult to obtain. For example, in cooperative group situations, how will standard conditions and standard criteria be possible? Will the evaluators be able to distinguish knowledge, basic skills, and performance? As to efficiency, feasibility concerns will determine the type, the quality, and the quantity of the assessment in a school environment. Finally, what will be the turn-around time between assessment and the teacher obtaining the results?

Given examples of current practice in performance assessment, in what sense have they worked well?

Some areas of the curriculum are already performance-based, e.g., physical education and vocational education. However, these activities do not go to the extent desired for authentic, engaging, and rich criteria. The approaches of literature-based language arts can lead to integrated thematic units in the curriculum. These current forms require a knowledgeable, skilled teacher, dedicated to the goals of the program. The teacher, therefore, must become a risk-taker, one willing to try new approaches in instruction and assessment.

Are there implications of performance assessment in terms of fairly assessing the different genders, racial and ethnic groups, and social classes?

In Kentucky, the Appalachian population is a major concern. Will the "language" utilized in the assessment be appropriate for this minority? There is the possibility that low assessments may represent the language unfamiliarity of the assessment, rather than low performance. This same concern can occur for the inner-city African American populations of the state. Students' environmental and cultural background will indeed have an impact on their performance. The assessment will need to make sure this impact is not a negative one.

Finally, there is the unresolved problem of the special needs populations. Mainstreaming students can fluctuate a classroom or a school's's assessment. Special logistics will need to be standardized for these students.

Do the activities defined, illustrated, and discussed meet the requirements of the law? If not, what must be done to create a fit?

For whole state feasibility, matrix sampling utilizing scoring rubrics will be required. School districts will need to derive their own assessments for non-fourth, eighth, and twelfth grades. Districts will thus have the responsibility, but not necessarily the rights of decision making, for performance-based assessment. Districts, in a sense, will need to "get in line" with the state assessment criteria and procedures for their own assessments. An appeals process will need to be in place for districts and schools to use if the institutions disagree with the state results or criteria.

What are the structural barriers, if any to using these activities in the schools? e.g., Are schools prepared to implement these new kinds of testing procedures and can they do it routinely?

After the assessment is changed, the next physical barrier will be the classroom textbook. Unfortunately, a majority of teachers tend to rely on their texts as a curriculum guides for instruction. The type of instruction envisioned by performance-based assessment differs from the skill-and-drill textbooks currently available. The monetary, political influence of the textbook industry may try to intervene in this area of school reform.

Another major barrier will be the teachers. Currently teachers are poorly informed of the exact details of the performance-based assessment they implement. Teachers, as well as administrators, will need to obtain "ownership" of the assessment if positive change is to occur.

What are the costs of implementing these new activities? (Materials as well as labor)

After state assessment instruments are developed, the major expense will be staff development. Teachers and administrators will need to be informed of, then included in, the implementation stage of the assessment. Inservices will need to be developed and presented where educations can develop, test, and evaluate their own types of per-

formance assessments so to be comfortable with this new area of education.

The reform act provides monies for the state assessment, but who will provide the cost of developing and implementing district-level assessments for the other grade levels? If continuous assessment is to be obtained, will not a continuous, on-going cost occur?

Are there areas in which training will be required? What areas and what training?

Following staff development for the development and implementation stages of the assessment, even more inservice assistance will be needed in the area of instruction. The curriculum performance-based assessment reinforces is much different than the current scope and sequence. Many teachers will be asked to instruct in a manner different than their current practice of lecture and individualized practice. Teachers cannot be expected to "reform" their practices without assistance and without support.

The national professional organizations, e.g., National Council of Teachers of Mathematics, already have taken a lead in the development of curriculum, instruction, and assessment methodology that can aid the classroom teacher. The problem is the inservices dealing with the state reform act are using up all available professional days and monies. Teachers are having difficulty attending and participating in the national organizations because of the financial and time requirements of KERA. Currently, the professional organizations are competing with KERA instead of complementing it.

What are the relationships, if any, among performance assessment, school-based decision making and other aspects of KERA?

The aspects of KERA are broad and wide-reaching. Education in Kentucky will need to change dramatically to conform to KERA's requirements. By changing the entire system, instead of a piecemeal approach, educators are now aware, even excited, about the reforms. So the first two steps toward attitude change are occurring, namely, awareness and commitment.

For substantial change to occur, educators and the public will need financial and logistical support. To this end, the various aspects of the reform act may impede each other. Just so much staff development and curriculum development can occur during a given time interval before all change is rejected. Careful watch must be kept to make certain all the reforms do not overload the education system nor the educators.

Summary

More than any aspect of the Kentucky Education Reform Act, performance-based assessment will change, hopefully reform, the curricula in the schools. To some teachers in the Commonwealth, the assessment will be a reward for the quality job they have been doing for a long time; to some, the assessment will be a punishment to force long overdue changes in school content and method; but to all educators, the assessment will direct the daily classroom instruction towards a more realistic, practical, applicable, and yet rigorous level. To the public, assessment is a way to obtain accountability of the use of precious resources -- our time, our money, our children; our future. Yet to many educators, assessment is viewed as a possible punishment -- it is feared. Teachers must be totally involved in the design, installation, and evaluation of performance-based assessment for it to be successful. How the new assessment in constructed, implemented, and then utilized will determine if performance-based assessment will obtain its potential as a beacon to guide education in Kentucky.

References

Kentucky Education Reform Act of 1990. (April 11, 1990). Kentucky House Bill 940. Frankfort, KY: Kentucky General Assembly.

Legislative Research Committee. (1990). A guide to the Kentucky Education reform act of 1990. Frankfort, KY: Kentucky General Assembly.

Mitchell, R. (1989). A sampler of authentic assessment: What it is and what it looks like. Paper presented at Beyond the Bubble: 1989 Curriculum/Assessment Alignment Conference. October, 1989: Sacramento, CA.

Pankratz, R. (1990). What is performance assessment? Unpublished paper for Council on School Performance Standards.

Wiggins, G. (1989). The futility of trying to teach everything of importance. Educational Leadership, 47 (3), 44-59.

Middle School

Bette Burruss

Rob Kennedy

Approximately 28 energetic school personnel from around the state gathered to learn new information about the Kentucky Education Reform Act of 1990. They were asked, "What does performance assessment mean to you? Their responses were in the form of statements that could be made and questions that still must be answered. These responses may be grouped as follows.

Broad definitions:

What do you know and how do you show it?

Demonstrating that a student can apply what has been learned.

Interdisciplinary testing across subject areas.

Correlation of goals and objectives and how well they are met.

Characteristics of performance assessment:

Performance assessment involves random sampling; practical applications; and multiple-choice questions.

Writing samples may be included.

Multiple instruments may be required.

Performance assessment is integrated across all areas of the curriculum.

It includes basic knowledge, process skills, and higher order thinking skills.

Performance assessment involves conceptual teaching and a focus on individual styles and needs.

Concerns about assessment:

Are students and teachers ready to move to performance-based assessment?

There may be more than one right answer to some test questions.

Requires a change in evaluation procedures.

Caution is needed.

Special populations, such as the gifted, will need special consideration.

There is a need to lay the groundwork for teachers. They may need training if they are unfamiliar with this method of assessment.

Performance assessment involves the evaluation of the school rather than the individual.

There are many questions now, but few answers.

Are we going to assess ability or what was taught?

There is a problem in comparing schools.

Is this just another reform wave?

How will local assessment be affected?

It seems like we are painting an old house with new paint.

There is an appearance of two testing systems--one NAEP and one performance-based. Are these the same?

Cost is a concern.

How will these tests be used?

How will feedback from the tests be used?

It will require a change in teaching emphasis.

We will need to help teachers with it.

Additional work will be required beyond what we are already burdened with.

There will be an effect on the dynamics of the classroom.

There are questions about the threshold. Should we start low so we can be assured of progress?

We will still need a measure of individual performance, especially for parents. For example, the CTBS may be used.

The ability range of students is wide.

There is a lot to do in a small amount of time.

Grading turnaround will be too long. For example, NAEP tests require a year to be graded.

The afternoon session included 16 diverse school personnel also in search of answers about student performance assessment under KERA. They were first asked for a working example of performance

assessment. Among the ideas shared were the following.

A scenario involving shoe sales entitled "Who wins the shoe war?" which could incorporate both mathematical and communication skills.

The debate over the right to bear arms addressing both the Constitution and communication.

A job application with connections to life, probability games, and self concept.

The use of the class as a student council for group dynamics.

Recycling as an application project.

Family economic awareness involving a budget and the cost of raising a family.

The group was next asked, **"What are the criteria for good performance assessment?** Their responses are listed as follows.

It should:
be interdisciplinary.
be a synthesis of the students' learning.
include conceptualization and application.
be reliable.
be clear.
be continuous.
be appropriate.
be facilitative.
be fun.
include cooperative learning.
be creative.

The third question posed the group was **"What are examples of current practices in this area which have worked well?"** Their responses included the following:

The Writing Project
Portfolios including writing samples and holistic scoring which can be passed on to other teachers in subsequent grades.
Manipulative math
Hands-on science
Science-math integrated projects

The fourth question was **"What new information is needed relative to student performance assessment?"** Their answers included:

People can get excited about student performance assessment!

We're all in it together and can learn together what shape it will take.

There is an opportunity for ownership from teachers.

There will be a need for teacher inservice activities.

How will it change the schools and students?

Are we a sounding board?

We are asking questions but not receiving answers.

There will be a lot of power in the hands of textbook publishers.

What is the relationship of performance assessment to other areas?

There are many different ideas about performance assessment.

There is a need for what the questions are going to look like, for example, in math.

There are a number of test issues, for example, problems with validity and reliability.

There is a need for input from practitioners.

Higher education is enthusiastic about helping direct assessment.

There is a need to continue doing what we are already doing right.

We need to involve more teachers.

Finally, the group was asked to identify **barriers and facilitators related to student performance assessment.** Among the barriers listed were:

Mind set of seasoned teachers
Validity of performance based assessment
Who are the evaluators?
Who's in control?
Time to do the assessment
Time, people, and money
Fear, accountability
Geographic bias
Setting parameters so can develop the curriculum
Funding
Leveling off point

Interdisciplinary curriculum planning, staff development, and instructional strategies were viewed as facilitators.

Reading, Writing and Language Arts

Peter Winograd

Jim Craig

What is performance assessment in reading and writing? What are the criteria for deciding whether they are good or bad? Are there workable examples of performance assessments in the language arts? How can we implement performance assessments in the language arts in Kentucky's schools? What kinds of training will be required? These and other questions provided the impetus for spirited discussion and debates among the teachers, administrators, and university professors who attended our sessions.

We came to some conclusions during our discussions, but we raised more questions than we answered. Perhaps the most important lesson learned was that many of our concerns were shared by others. We will present the results of our deliberations in a series of interrelated points.

1. Performance assessment in reading, writing, and the other language arts must focus on important outcomes rather than on fragmented or isolated skills. Some of the important outcomes identified by our discussion groups included the ability to comprehend what has been read; the ability to write clearly for a variety of authentic purposes; a love of reading; the ability to use reading and writing to grow in affective, personal ways; and the ability to apply reading, writing, and the language arts effectively across a wide variety of subject areas.

2. A great deal of our discussion focused on what it means to comprehend. We believe students should be able to read authentic texts and, depending on the type of text and the purpose for reading, display a global understanding, a more detailed understanding, a personal response, or a critical evaluation.

3. Participants in our sessions expressed their hopes that writing would become more common in many subject areas and not just in English. For example, writing in response to what is read in social studies or science is as valid as performance assessment task as is writing in response to what is read in literature. This statement, along with points 1 and 2, reflects the importance of ensuring that the language arts are integrated with other subject areas.

4. Performance assessment in reading, writing, and the language arts should include authentic texts and entail a variety of response formats for students. Students can display their understanding in a number of ways including writing in response to reading; discussing what they have read; working in groups to extend and communicate their understanding of what they have read; following or composing directions; role-playing their response; and drawing, using graphic organizers, or other visual aids.

5. Performance assessment in the language arts should include some tasks that students do on their own and some tasks that require cooperative efforts.

6. Performance assessment should be part of long-term instruction, not something that happens once in a while. In order to accomplish this integration in a constructive manner both assessment tasks and instructional tasks must focus on authentic activities. District level scope and sequence guidelines for the language arts should be updated to reflect insights from current research and the requirements of KERA.

7. A tremendous amount of training and support needs to be provided to Kentucky's teachers. The desired improvements in assessment and instruction envisioned by KERA will take more time and effort from teachers and administrators who are struggling to meet current demands. For example, if students are expected to engage in more writing, then teachers will need additional time and training so

they can respond in ways that help their students grow as writers. In addition, teachers and administrators will need additional assistance in helping students, parents, and the community understand the changes that are made in how we assess growth in the language arts. In addition to these points, the discussion groups thought the following examples were useful in illustrating workable performance assessment tasks in the language arts. These performance assessment tasks focus on reading and writing for understanding, and reading for pleasure. These are suggestions for students in grade 4; the tasks would be adjusted for younger or older students.

Language Arts Goal: Reads to understand a variety of print materials for a variety of purposes. Writes to communicate ideas to a variety of audiences.

Performance Assessment Task

Writing in Response to Reading. The student will read two different selections on a common topic. Possible examples might include two different editorials dealing with the same issue (e.g., crime control, pollution legislation); two different authors describing the same event (e.g., growing up in the South); or two different authors describing historical events (e.g., a Northern and Southern view of the Civil War).

The student will then write a response that analyzes and evaluates the two different reading selections. In addition the student is able to communicate his or her own reaction to, or view of the issue or topic under consideration.

Potential Questions could include the following:

1. Which author presents a more convincing argument? Why?
2. Could the arguments presented by both authors be valid? If so, how? If not, why not?
3. How is your view of the topic different from or similar to the views presented by these two authors?
4. How do you think these two authors would respond to your interpretation or point of view?

Performance Criteria

Possible Performance Criteria could include the following:

1. Comprehension - the student is able to understand the gist of the selections he or she read.
2. Analytical/Critical Stance - the student is able to compare and contrast the selections in terms of the quality of the arguments.
3. Personal Response - the student is able to articulate his or her own points of view and compare and contrast this view with those expressed by the two different authors.

Scoring Rubric

Expert Performance

The student displays an excellent understanding of both selections. He or she is able to compare and contrast the writing of the two different authors, and he or she is able to develop and defend his or her own interpretation.

Acceptable Performance

The student displays some understanding of both articles. He or she is able to compare and contrast the two selections, but only along a relatively few dimensions. The student's personal interpretation is relatively short and without much substance.

Sub-standard Performance

The student displays little understanding of either article. His or her own interpretation of the issue or topic under consideration is weak.

Language Arts Goal: Reading is a major source of personal enjoyment, inspiration, and growth.

Performance Assessment Task

1. Ask the student to bring up a book selected from his or her list of books and magazines read for pleasure (e.g. during free reading at school).
2. Have the student talk about the book. Possible questions include: a. Why did you pick this book to read? b. Tell me about this book? c. Would you recommend this book to a friend?
3. Ask the student about others books on his or her list. Possible questions include: a. Which authors do you like the best? b. Why do you like their books? c. What kinds of things do you like to read about?

Performance Criteria

Several criteria could be applied to this performance task, but three of the most important follow:

1. Enthusiasm or other evidence that the student views reading as relevant to his or her personal life.

2. Evidence that the student understood the gist of the book or experienced a strong personal response to something he or she read.

3. Evidence that the student engages in reading for pleasure on a regular basis.

Scoring Rubric

Expert Performance

The student has an extensive list of books and magazines read for pleasure. The student's description of the book reveals an appreciation of the author's style. The student reveals some insightful understanding of how the book is relevant to his or her own life.

Acceptable Performance

The student's list of books is relatively short, perhaps the student is reading only short, familiar books. The student knows a limited number of authors. The student is able to remember some things about the book, but exhibits relatively little enthusiasm or understanding.

Sub-standard Performance

The student's list of books read for pleasure is extremely short or non-existent. The student has difficulty identifying favorite stories or authors.

The Practical Arts

Sandra Miller

The following points were made/discussed at the Practical Arts session:

1. The goals, philosophy and traditions of the Practical Arts lend themselves more readily to performance assessment than do some other subject areas.

2. The Assessment Component of the Kentucky Competency Based Vocational Education (CBVE) Curriculum is performance based. CBVE is based on validated tasks which are specified in measurable behavioral objectives.

3. Physical Best is an illustration of an effective assessment based program. Each student identifies his/her physical limitations based on national standards and then sets goals. Periodically students assess progress toward accomplishing the goals, establishing new goals as appropriate.

4. Performance assessment will be labor intensive. In the Practical Arts it will also be expensive in terms of products and equipment.

5. A problem in assessing students in any subject area is that they bring their deficits in other subjects to the subject under review.

6. Performance based assessment instruction might be organized by learning tasks that require interdisciplinary instruction. Secondary teachers will find interdisciplinary teaching more challenging than will elementary teachers.

7. The Skyline Vocational Center in Dallas, Texas has an impressive system of keeping records on student performance.

8. The Kentucky Education Reform Act specified that all students will be successful.

- It will be difficult to define success, since what is successful in one person is not necessarily considered successful in someone else.
- Since a person's ability to make a living is often considered a sign of success, business and industry ought to have input into the definition of success.
- In order for students to be successful, they will need to be guided into classes compatible with their needs, aptitudes, and interests.
- Students could be motivated to be successful by incentive programs in which they earn items (e.g., badges) based on level of proficiency in learning task completion. Youth organizations which have proficiency incentive programs include Boy Scouts, Girl Scouts, and various vocational education student organizations.*

9. KERA requires measurement of the proportion of students making a successful transition to work, postsecondary education, and the military. Criteria to be used in measuring success is of concern as is the manner in which data will be collected.

10. Training via satellite transmission may be an efficient way to prepare educators to do performance assessment.

11. A new position that may evolve in local districts is assessment specialist. University programs may need to be developed to prepare people for these jobs.

12. School based decision making groups will need to be well versed in performance assessment. Further, they will need good data on which to make decisions. It will be incumbent on districts to provide the best data possible in order for groups to make the most informed decisions.

13. In order to develop a truly effective performance assessment system for Kentucky, its purposes will have to be decided and clearly stated.

*Vocational education was assumed to be a Practical Art since no other workshop at the Conference encompassed it.

BEST COPY AVAILABLE

Arts and Humanities

Ronald Schmelzer

George Szekely

As defined in our discussion, performance based assessment is the student demonstrating his or her skills, concept attainment and/or creativity in the arts or humanities. These skills, ideas that have been learned and creative performance, may be evaluated against three criteria that are relative to: peers, personal growth, and a predetermined objective. A student's performance may be judged as Basic (having achieved minimal competency); Proficient; and Advanced.

Participants agreed that in the performing arts, e.g., music, visual arts, dance and drama, emphasis should be placed on the creative aspects of the art. This is not to say that skills and concept development should be slighted; although in some instances, this may legitimately occur.

It was generally agreed that more questions had been raised than answers given. Some of the questions raised follow:

1. How much responsibility and freedom does the individual teacher have in developing his or her own criteria for performance?
2. What role does the students' past experience in the arts have in performance assessment? In other words, should a student who has a good deal of competence be judged by the same criteria as one having no competence?
3. Is there a common cannon for the arts? If so, what role will it have in creating rubrics for assessment?
4. What role should professional associations have in developing performance objectives?
5. Should the same criteria be applied to different regions of the state? different subcultures?
6. What effect does the quality of instruction have on the attainment of performance standards? If performance is influenced by instruction should all teachers be equally competent? How will this be assured?

In the afternoon session there was much discussion and little agreement. Some of the comments and questions follows:

1. What is performance based assessment?
2. Performance based assessment has been ill-defined.
3. Little to no guidance has been given as to how to implement.
4. Teachers are once again being given the responsibility to implement something, but not the freedom or authority.
5. An overall structure needs to be developed and communicated that will allow for the development of a reasonable performance based assessment plan.
6. Politicians and administrators need to give the teachers time outside of the classroom to develop performance based assessment.
7. A consensus needs to be developed in regard to the value of performance based assessment.
8. Many, if not most, administrators fear that their authority will be usurped by site based management and performance based assessment. A consequence of this is that it is being "bad-mouthed."
9. Staff development is critical to the development of performance based assessment.
10. Teachers need to be provided an environment that is conducive to collaboration, including common planning time. Unless this happens, there will a need for additional staff; indeed, this may be the case anyway.
11. Constraints place upon a teacher's instructional time must be lifted.
12. Colleges and universities must take the lead in guiding school districts in the development and implementation of performance based assessment.

Group Leaders' Comments:

The group experienced a great deal of anger and frustration during the session. This may have occurred because few, if any, of the participants had an understanding of what performance based assessment is, much less how to implement it.

We believe that there is a critical need for staff development on a state-wide basis in order to clear up misunderstanding and to help teachers and administrators recognize the value of performance based assessment including the critical need for site-based management as integral to its implementation.

Reform Law

Virginia Davis Nordin

Joseph M. Petrosko

The Kentucky Educational Reform Act (KERA) is a vaguely written law. Compared to other educational laws, it has many undefined terms and many areas of ambiguity. The intentional vagueness of KERA presents both problems and opportunities. One problem might be an excessive concern for performance outcomes. Schooling might become totally redirected to the achievement of observable outcomes and nothing else. However, the vagueness of KERA provides opportunities to educators. For example, there may be a heightened awareness of the needs of deprived students. Issues of economic equity will become obvious when both rich and poor students start being held accountable to the same performance standards.

A problem with performance assessment (as it has evolved so far) is lack of local ownership of the process. Teachers and school personnel do not have a feeling of participation. Various task forces of the Council on School Performance Standards, for example, task forces in mathematics and language arts, do involve substantial numbers of teachers. However, the great majority of Kentucky teachers are merely onlookers to the process of writing performance assessment tasks. Better communication links to the local level are needed.

Comparability and accountability are key issues in the Kentucky school reform. Each school will be held accountable for the achievement of its students, but how fair is this when schools are not comparable? The law recognizes that not all schools are the same. Nevertheless, the manner in which different schools will be treated is presently unknown.

Local control of schools, in the form of site-based management, is relevant to the discussion about performance assessment. According to the law, each local school council will adopt a policy regarding "determination of curriculum. . .curriculum development. . .[and] alignment with state standards." Does this mean that local councils set goals for achievement? If so, will there be a procedure to assure that there is comparability among schools in a district? Furthermore, what about goals coming from the state department of education? Model curricula for elementary and secondary education are supposed to be published in 1993. But during 1991-1992, interim testing will have already occurred to establish a baseline for assessment decisions to be made during the 1993-1994 school year. So what is the point of the state's model curricula? It seems that schools are held responsible in the assessment procedure, but with little control over or understanding of standards.

The method of reporting results of performance testing will be very important. State assessment is designed to focus on the individual school. However, the news media are sure to take the published data from state assessment and attempt to rank districts, just as was done with the Kentucky Essentials Skills Test (KEST) and Comprehensive Tests of Basic Skills (CTBS). This will probably have a negative effect on some districts. How can this be avoided?

The relevance of performance testing to parents is questionable. Many parents are solely concerned with how their child performs and compares with others. They are not concerned with how the school performs. Results of traditional testing--the continuous assessment program at the district level--may be of more interest to parents than results of performance testing.

The exact relationship between performance testing and continuous assessment program testing has not been worked out. For example, it is conceivable that students in a school could do poorly on performance testing, but do well on a standardized test used in its district's continuous assessment program. Such a discrepancy could undermine the credibility of rewards or sanctions given to the school.

According to KERA, schools will receive rewards or sanctions depending on changes in the percentage of successful students. The precise meaning of successful student is currently unknown, and will require a lot of effort to specify. The state department of education has produced a table, a "Reward Matrix," which lists the consequences of each outcome (e.g., what happens when the percentage of successful students decreases by more than 5%). Most educators believe that this has an air of unreality. There are too many unknowns. Too many things have to transpire between now and the decision dates shown on the matrix to take these various consequences seriously.

Someone should be taking note of the unintended consequences of KERA. For example, in implementing performance assessment, we may end up with students who perform competently during a testing situation, but who never develop the disposition to perform well on their own. They will be good test takers, but not very good life-long learners.

Assessment tasks need to be relevant to diverse populations. An example of an assessment task proposed for use in the schools requires students to prepare an oral report on the likelihood of an earthquake in Kentucky. The purpose of the report is to help the mother or father of a student to decide on the purchase of earthquake insurance. In some schools, many students do not live with either parent. Also, students do not live in a house, but rather an apartment or housing project. Finally, many students do not know what insurance is because their families do not have it.

Afternoon session

Educators should learn about the concept of technical amendment to a law. The KERA of 1990 is sure to require such amendments. They could be put into a bill that would be voted on by the General Assembly. Technical amendments would clarify and modify ambiguous portions of the law. They would not change the essential nature of the law, only refine it. Educators should think about proposing specific, practical technical amendments.

A number of things in KERA could stand modification. For example, the law requires performance assessment at grades 4,8, and 12. It might make more sense to begin testing at grade 3 rather than grade 4 (especially in light of the shift to ungraded early primary school). Further, the relationship between assessments required under Sections 2 and 3 of KERA needs clarifying, as well as the relationship between curriculum and assessment responsibility.

Although the vagueness of the law allows educators flexibility, the strong sanctions for "failure" (the latter essentially undefined) seems unfair and seems to lead to unnecessary and nonproductive anxieties.

At the present time there are no legal requirements in effect except those for continuous testing. The requirements will arise only after the Council on School Performance Standards and other groups issue various standards, guidelines, and regulations.

The costs of implementing performance assessment will be substantial. A lot of time and money will be needed to train teachers. There will also need to be a system for sending out material from one school district to another. Biases could occur if teachers from a school tried to assess their own students.

Implementation and Management

Kawanna Simpson

Scott Trimble

Ken Draut

Tom Mowery

The purpose of the session was to give school leaders a better idea of their responsibilities to manage the new testing program. The process was for Draut and Trimble to present important information for 30 minutes followed by group discussion of several issues.

What has student assessment in Kentucky looked like in the past?

Trimble gave a historical sketch of Kentucky mandated testing program:

- mid 1960's the state began helping to score state tests to be used for standard comparisons on a voluntary basis.
- early 1970's a need grew for regional and state testing data.
- 1978 the state testing program changed focus with the Education Improvement Act that required testing in grades 3, 5, 7 and 10.
- 1984 the KEST was mandated that increased the level of accountability--also provided estimated norms.
- 1988 the legislature mandated a norm-referenced test with no estimated norms.
- brings us to 1990 KERA and a whole new testing program.

What went wrong with past assessment programs?

The problems in Kentucky were not unlike national problems:

- Due to rankings of school districts, immense pressure was placed on norm-referenced tests.
- The samples of curriculum on the tests frequently became **the curriculum**--too much focus on items rather than the whole curriculum.
- The curriculum became too narrow, which tampered with the validity of the tests.
- The norms could not be updated annually, leading to most states placing above the national mean.

What is performance assessment?

Draut led a discussion of the session materials which included samples of performance assessment from other states and from Kentucky sources. Many teachers have for years routinely included similar activities in their curriculums.

There will be pressure for teachers to change the way they teach so that methods reflect the new ways of assessing student performance.

How are performance assessment items scored?

The scoring of performance based assessment items will generally be more complex than past state testing programs.

Much of the scoring will occur at the school level rather than sending answer sheets to Frankfort for scanning and scoring.

Teachers and others must be trained to judge student performance in ways foreign to many. As examples of scoring difficulty, OM and Governors Cup Future Problem Solving competitions were noted. Many educators are aware of the difficulty in training officials for these competitions, and the large intervals of time required to judge them.

Question: Who will train the judges? When? Where? Cost?

Training must be a joint effort of higher education, Department of Education, local schools and others. Most of the cost will probably be paid by local school districts. KET must have a key place in the training of judges.

Question: Will additional in-service days be provided for teachers and administrators to become trained?

This is a concern that must be addressed by the legislature, but training is a priority.

Question: Who will be trained to write the assessment items that frequently require much attention to detail so that students know exactly what is required of them?

Question: Proper judging (i.e., reliable judging) is essential to prevent lawsuits since sanctions and rewards will be based on the outcomes. Are we underestimating the training of judges?

Question: Isn't the only true way to have reliability is to use the same items over and over across all school buildings?

What are some major issues in the state performance assessment program?

- Sampling, reliability and validity are the three issues.
- Design of the test will have to take these issues into account.

What are some examples of current statewide assessment programs?

Draut addressed what some other states are doing. Arizona in particular has developed an interesting system. They develop parallel tests, one to be given for teacher use early in the school year, and the second to be given in the spring semester for assessment purposes. In addition, Arizona gives a multiple choice norm-referenced test each fall semester to meet the demands of those who want such information.

How will performance assessment impact local school assessment?

- Local districts must have a plan for continuous assessment that goes beyond the state program. It seems reasonable that schools will continue to give traditional tests to meet local information needs, such as determining basic skill achievement, communicate with the community, and to have Chapter I data.
- Schools should also give tests similar to the state assessment to students in grades other than the three grades mandated by the state program.
- To have a quality local program, districts need to provide staff training using writing prompts and other prompts.

Question: What should a school do first?

Vocational and special education teachers have been using performance assessment for some time to measure outcomes. They might be a good first source for ideas.

Be thinking about how technology can be used to help with performance assessment, both in the assessment phase and the judging phase.

Question: Who determines the mastery level of a task?

Until the state gives some indication, schools should be experimenting with mastery levels.

Question: Will the Department score CTBS tests this spring?

They will score them this year and perhaps every year, depending upon the new commissioner and new state board.

Question: What can schools be doing in the next six months?

There are only best guesses right now about the new testing program. All schools should begin experimenting and practicing. Some national

groups are working with mastery learning, so teachers should get active in their national and state associations. Get materials from OM and Future Problem Solving competitions and scouting merit badge books. These sources of performances are handy even though rubrics are not given in all cases to judge the performances.

Question: What should schools do to prepare students for ACT and SAT tests for college entrance?

Scores on these tests will continue to be a measure of school program success. Schools should continue to have multiple choice tests as a part of their continuous testing program.

Question: Is the Department of Education still going to offer the statewide in-service on student performance assessment on February 15?

Yes.

Question: With all the demands of KERA on local school staff, how can they find the time to train in all the areas?

This is a concern that must be addressed quickly by by proper state authorities, but it could be that local school officials will have to find a way for training at the local level.

Question: How will the School Performance Standards Council impact the performance assessment program?

There is a close working relationship between the two groups.

Question: Since the state did not pay $300 for positive teacher evaluations under a 1985 state program, how do teachers know that money will be available for rewards and also to provide service to schools in crisis?

There is never any certainty that a program will be entirely funded, but the tax changes and legislative commitment seem to indicate funds will be available.

Question: How will the state inhibit cheating in the new program?

A new code of ethics is needed.

It could be that there will be testing centers independent of the Department or schools to which students go to be tested. Or, teams of educators from one district could go to another district to administer and judge the performance assessment. Outside monitors or administrators may be used.

What are the barriers, if any, to implementing performance assessment? materials? costs? training?

The costs of the program have probably been underestimated, the materials will have to be developed, and the training will be costly in dollars and time.

How will school based decision making and performance assessment impact each other?

- Teachers at each school will have rewards and sanctions impacting their lives due to performance assessment. They have, then, a vested interest in how their students perform.
- Many teachers will want more say in the treatment of students since their careers might be at stake.

Discussion questions:

1. How will the following elements of reform impact performance assessment:

   - Accountability (rewards and sanctions)
   - Ungraded Primary
   - Special Education
   - Administrative and Instructional Technology
   - Preschool Programs

2. What management problems can be predicted when implementing performance assessment in elementary, middle and high schools?

3. How can the state pay for rewards for schools and also provide service to schools in crisis?

4. What changes in information needs will be required due to performance assessment?

5. What type of materials should schools be developing in preparation for performance assessment?

Additional questions from the session participants:

Question: Will the Department provide guidelines for Chapter I testing?

Yes, but these generally come from federal mandates.

Question: What areas will be assessed?

At least the six goal areas listed in the Reform Act will be tested.

Question: Should teachers teach the test?

It appears that the intent of performance assessment is to change the way teachers teach. It is hoped that the way teachers teach will become natural preparation for the tests. In fact, teachers should use performance assessment routinely to grade students. Teaching to the test will be a positive experience for teachers and students.

Question: How can we prevent the Lexington Herald Leader newspaper from ranking schools?

Cannot. In fact, schools may start being ranked.

Question: How is a school's threshold changed when a drastic population shift occurs that might significantly increase or lower the threshold?

A formula must be developed to account for such shifts.

Question: How will this program provide national comparisons for Kentucky schools?

And some other Questions:

Question: How will this program make better teachers?

Question: If we have not been successful in teaching lower order thinking skills, what makes anyone think that we'll do better with higher order thinking skills?

Question: Why were 12th graders chosen to take the exam, given their likely uncaring attitudes?

Question: When will colleges catch up to performance assessment?

Question: Where is student accountability in this assessment program?

Question: How will public schools compete with private schools given this experimental program?

Question: Since multiple choice tests take less time, are objective, are easy to score, and many questions can be covered, what are the advantages of performance assessment?

Question: What if a school threshold is so close to the top of possibilities that they can't easily get rewards?

Question: When must a LEA have a continuous assessment program in place?

Question: Must a school district include kindergarten in their local assessment program?

Question: Will the Department continue its test item bank? When will it have performance assessment items in it?

Question: How will performance assessment help determine if a child is ready to graduate to the fourth grade?

Item Sampling

George Cunningham

Skip Kifer

Matrix or item sampling is a set of measurement techniques and linked statistical methodologies which under certain circumstances allows one to save time, money (or both) and increase the amount of information gathered when implementing a testing program. These procedures are used extensively in the National Assessment of Educational Program (NAEP) and state-wide assessment programs.

Consider, for instance, the problem of estimating the percent of twelfth graders in the Commonwealth who can respond at a satisfactory level to the question:

Why in the United States are slave insurrections called revolts, not wars?

One approach to estimating the proportion of students who answered at a satisfactory level would be to give the question to all students in grade 12 and then calculate the proportion who gave satisfactory answers. That would mean measuring thousands of students in about 300 high schools. This method produces precise, valid, and useable results, but is not efficient in a statistical sense. Comparable results could be gotten from responses of many fewer students. (The Gallup poll estimates percents of the U.S. public on sample sizes around 1000.)

Approaches using item or matrix sampling could change the above design in two major ways. First, because there are so many 12th grade students, one could define four tasks (instead of one) and randomly assign one those tasks to each student within each of the schools. One could now estimate four proportions (albeit with slightly less precision) for about the same amount of time and money using the same population that provides just one estimate in the first scenario. Hence, through the sampling of tasks, one has increased the amount of information gathered per unit of time and money.

Second, one could take a sample of schools and administer just one task to all twelfth graders in the school. That sample would be used to estimate the proportion of students within the Commonwealth who gave satisfactory answers to the question. Hence, through the sampling of units, one has saved money (and perhaps time) but still obtained the desired estimate.

A combination of the two major ways of sampling (units and tasks or matrix sampling) provides a more efficient way still to gather information. This time one could sample schools and give samples of tasks. For example, one might define eight tasks and give a different subset of the eight to a sample of schools. In this scenario one is sampling both tasks and schools and is, theoretically saving money (perhaps time) and getting increased information about a broader array of tasks. This kind of sampling can be expanded so long as there are large enough samples of students taking each task to estimate precisely a proportion of interest.

The notions of matrix or item sampling figure prominently in the language KERA. No distinction is made, however, among task, unit or matrix sampling. It is clear that a *census* (the first example where all students do all things) of students and tasks is not required. The language does, however, allow for the possibility of *task* sampling, *unit* sampling and both (*matrix* sampling). To the extent the law requires a state-wide assessment, a form of sampling is a way to get desired information for less cost and time than by having all students respond to all questions. In this sense item sampling would be used in the Commonwealth just as it has been used in NAEP or other state-wide assessments.

A second use of item sampling was apparently envisaged by those who framed the law. The accountability portion of the legislation makes a school the unit which could receive rewards or sanctions. Item sampling, theoretically, would allow one to estimate a percent of successful students in a school

without requiring all students to do all tasks. Hence, item sampling is viewed as a cost effective way to implement the accountability arm of the legislation.

It is in this second use that the intent of the legislation may be pushing the psychometric limits of item sampling technology. Three features of the law or Kentucky's schools should be noted in this regard:

- There are small schools in Kentucky and there are schools with substantial transient populations of students.
- Comparisons to be made are between cohorts. That is, performance in 1992 for grade 4 students is to be compared to performance of students in grade 4 in 1994.
- The precision with which one must estimate the percent of successful students is less than one percent. That is, a school that is to receive a reward must do one percent better than its threshold value.

These features, individually and in combination, place severe pressure on this set of methodologies. The premise of item sampling is a large population of things (in this case students) which can be sampled without a great loss of precision when making statistical estimates. Suppose, however, that a school had just fifty 4th graders. If each student were to do one task, a response would be "worth" two percent. If each student were to do one of two tasks, the response would now be worth four percent. The implication is clear. Each student would need to do a lot of tasks in order to estimate with the desired precision.

Suppose two years latter there was a marginal change in the cohort. This new class had two students in it who could do everything well while the class that formed the baseline had only one such student. That one additional successful student could have a profound effect on the estimate of the percentage of successful students, the comparison between levels of that percent, and whether or not a school received a reward.

This example is meant to be just that. The problems of estimating the percent of successful students is difficult regardless of the size of school and compounded with cohort comparisons and the need for extremely precise estimates of a school's performance.

It may be the case that *matrix* sampling will be of limited utility in the accountability portion of KERA. A *census* of students with some tasks (each student does a number of common tasks) or task sampling may be more appropriate.

Performance Assessment for Special Populations

Deborah Bott Slaton

Jay McLaughlin

Teachers, administrators, and teacher educators attending the conference on performance assessment and the session dealing with special populations expressed several concerns regarding how students identified as handicapped or eligible for Chapter I services would take part in the performance assessment program developed for the Kentucky Educational Reform Act. Participants also noted some similarities between procedures for performance assessments and some informal assessment techniques used to monitor progress of some students with special needs. Finally, participants discussed existing assessment requirements for federally funded programs such as Chapter I and programs for students with identified disabilities.

Concerns About Participation

Session participants expressed concerns regarding how students with disabilities would be included or excluded from a state-wide performance assessment program. In some instances, the educational programs of students in special education will not match the content of assessment instruments designed for the non-handicapped school population. For example, students with moderate or severe mental handicaps would typically be working in a program that emphasizes functional, daily living skills rather than traditional academics. In other instances, specifically students with mild disabilities who spend only a portion of the school day away from the regular classroom, the content of educational programs of students with disabilities may match that of the non-handicapped school population although the match may not be with age-mates. Including special education students who spend time in the regular class in the performance assessment program could increase classroom teachers' sense of accountability for the rate of learning among students with disabilities. This was viewed as a potentially positive outcome of performance assessment.

Participants did not recommend that all students identified as handicapped be exempted from the performance assessment program. Rather, the consensus was that the Individualized Educational Plans (IEPs) required by federal regulation could be used to determine participation/exemption in performance assessment activities. This option would allow educators and parents to make decisions regarding participation on an individual basis, but teams would need guidelines and criteria for how to approach this decision. One possible component of the decision could be the amount of overlap found between curricula in the mainstream and the student's IEP goals and objectives.

Participants voiced concerns about how the rewards and sanctions tied to outcomes of performance assessment could pressure some schools with high numbers of students with special needs to automatically exclude these students from any performance assessment program. There were also questions regarding how special day schools and residential schools for students with disabilities (e.g., Kentucky School for the Deaf) would fit into the performance assessment program. Careful sampling procedures could control that the proportion of students with special needs present within testing sample is representative of the school or district student population.

Existing Assessment Practices in Special Programs

In some instances, existing assessment practices within special education programs may parallel or exemplify performance based assessments. For example, assessments of daily living skills using an ecological approach and community catalogs often require students to perform real tasks in community settings (e.g., using a city transportation system, shopping for groceries, etc.). Some of the curriculum-based assessment techniques used to assess IEP objectives may also be similar to performance-based assessments, depending on the nature of the

IEP objective. It may be possible for some assessments already conducted by special educators to be supplementary to and supportive of the performance assessment program. These examples of similarities among assessment procedures will not eliminate the need to include teachers working with special populations in staff development activities related to performance-based assessment techniques.

Existing Assessment Requirements in Programs Receiving Federal Funding

Several of the federally funded programs for students with special needs (i.e., Special Education and Chapter I) must comply with federal regulations regarding assessment, and the performance assessment program may not be an acceptable substitution for existing practices. Examples include the following:

- assessments required to establish presence of a disability and eligibility for special education
- ongoing monitoring and annual review of IEP goals for each student placed in special education
- screening of all primary students for eligibility for Chapter I services
- d) spring-to-spring administrations of norm-referenced tests required for students placed in Chapter I.

Participants did not foresee that the performance assessment program established under KERA will satisfy all program needs for data to modify instruction, inform parents, or identify which students are eligible for special services. Participants also raised concerns regarding confidentiality requirements for records of students identified as handicapped and the potential logistical problems involved if several different layers of assessment are required by federal and state programs.

Federal mandates for nondiscriminatory testing and reasonable accommodation must be part of any performance-based assessment program that involves individuals with identified disabilities. Again, the team writing the IEP is probably best suited to specify what accommodations an individual student would need to participate in a performance assessment program. Examples of testing accommodations include extra time, braille writers, and use of audio tapes to substitute for printed materials. The concept of "reasonable accommodations" should be used to guarantee that, when appropriate, students with disabilities participate fully in the performance assessment program.

Summary

There was a general consensus among session participants that students with special needs were not considered during the development of the Kentucky Educational Reform Act. The task before us now is to develop clear statements regarding how students served by special programs will take part in all aspects of KERA including performance assessment programs.

Student Services

Paul de Mesquita

Malessa Billings

This paper summarizes issues, viewpoints, and questions of a sample of Kentucky public school pupil services personnel who participated in group discussions focusing on the implementation of the Performance-Based Assessment section of the Kentucky Education Reform Act (K.E.R.A.). The group discussions were conducted as part of a series of workshops on performance assessment held at the University of Kentucky, January 4, 1991. A total of 26 discussants attended, representing superintendents, supervisors, principals, counselors, parents, school board members, and university faculty. Throughout the discussion, many of the issues raised revolved around four central themes: background and definitional questions, roles and responsibilities, implementation concerns, and future direction. This summary has been organized accordingly to reflect these four major themes.

1. Background & Definition

What is the purpose of performance-based assessment? Before a meaningful discussion of performance-based assessment could occur, some review of the background and context surrounding these issues was required. The state-wide educational reform plan for performance-based assessment is still relatively new and unfamiliar to most of Kentucky's educators. Consequently, many educators misunderstand the intent of such an assessment effort. The primary purpose of establishing a performance-based assessment program was to "insure school accountability for student achievement" of educational goals outlined in the reform act. A secondary purpose of the assessment program is to present a basis for comparing the performance of Kentucky students with students from other states across the nation.

What will be assessed? Beyond the goals and stated purposes of performance-bases assessment, another area of misunderstanding involves the focus of the assessment. The Kentucky Education Reform Act clearly delineates seven capacities that serve as the educational outcomes for the Commonwealth's system of public education. These broad areas include communication skills, knowledge and understanding of political and governmental processes, mental and physical well-being, cultural and historical appreciation, work skills, and skills to compete academically. The seven broad capacity areas will be translated into measurable goals and student outcomes that will serve as the basis for assessment of students' performance. Specific student outcomes will include not only achievement in traditional core subjects, but emphasize self-sufficiency, personal responsibility, thinking and problem-solving skills. Given the legislative purpose of performance-bases assessment and the stated outcomes that are to be assessed it is clear that performance-based assessment will have little resemblance to the more traditional standardized educational achievement test relied upon so heavily in the past.

What performance-based assessment is not.

Fully understanding performance assessment also means understanding what it is not. Performance-based assessment will not be a standardized testing program for all students. Typical paper and pencil, multiple-choice questions will not be the predominant format. Performance assessments will be more open-ended. Instead of filling in answers to questions, students may actively create solutions to problems. Not all children will participate. Only a representative sample of students will participate in the assessment process. Performance-based assessment will not yield the kinds of achievement score information that can be shared with parents of

individual students, such as a percentile rank. Keeping in mind the overall purpose of school accountability, the results will be most meaningful in terms of schools and school systems as a whole rather than measures of individual student achievement. The relationship between performance-based assessment as outlined in K.E.R.A. and annually administered standardized achievement testing remains unclear at this time. It is possible that these two different measurement perspectives eventually may be used in a compatible and coordinated way. This is an area that pupil services professionals may play a contributing role.

II. Role & Responsibilities

A second major theme of discussion focused upon the roles and responsibilities of student service personnel. A major area of responsibility probably will be that of inservice education for school personnel. Several possible areas in which leadership could be assumed are presented.

First, public school pupil services personnel must become knowledgeable of why students will be assessed and what kinds of student performance will be measured. This awareness and knowledge of the intent and scope of performance-based assessment becomes a responsibility for pupil services staff in terms of their role in educating colleagues, teachers, and parents about the proposed assessment program.

Second, since many pupil services staff function in positions such as guidance counselor, school psychologist, diagnostician, and supervisor, these professionals, because of their advanced graduate training, should be more knowledgeable about tests and measurement concepts. Therefore, they should be more informed about curriculum-based performance assessment as a more meaningful measurement approach. These views can be shared with school staff in ways that enhance the acceptability of the proposed performance-based assessment plan. Ways of integrating performance-based assessment with continuous classroom assessment strategies could be encouraged as a way of monitoring student progress through the curriculum.

Finally, many individuals in pupil services positions are sensitive to the stresses, frustrations, and related problems associated with the anticipated transition to educational reform in Kentucky schools. The development and implementation of performance-based assessment as a statewide school accountability system will bring added pressures to educators. Serving as a resource for information and guidance concerning performance-based assessment will be a major role for student services professionals. Helping educators understand their responsibilities and clarify the value of a performance-based approach to measuring student learning can contribute to maintaining positive attitudes and school morale.

III. Implementation Concerns

Throughout the discussion participants identified several potential barriers to the implementation of performance-based assessment. As alluded to earlier, lack of information and misunderstandings have created an atmosphere of pressure and frustration among many Kentucky educators. Relatedly, the continuing uncertainty surrounding how the assessment program will be developed and implemented encourages misinformation and increases anxiety on the part of schools who will be held accountable by such an unknown process. The fact that schools will be judged and evaluated by this as yet undetermined method has led to wide speculation about ways to circumvent the rules through teaching to the assessment procedures or overrepresenting high ability students in the testing sample.

Additional questions have arisen as to who will be responsible for administering the performance-based assessment within a given school or a given county. If school or county personnel are involved issues of bias and lack of objectivity may arise. It is possible that pupil services personnel may in fact be assigned to assist with administration or scoring the performance assessments. Knowing that unfavorable results may jeopardize their school's in terms of sanctions and rewards can create enormous pressures, not to mention ethical dilemmas.

From a more practical point of view several areas of concern were also voiced. These included how much time and effort would be required to administer and score the assessments? How much time would students be missing from class to participate in this process? At what point during the academic year would these assessments be conducted? What procedures would be used to insure that student samples would be random and balanced in terms of ethnicity, gender, and socioeconomic status. A considerable amount of inservice training related to performance-based assessment is necessary. Finally, what will the costs be in terms of

BEST COPY AVAILABLE

time, effort, and actual test development and will these costs be prohibitive?

IV. Recommendations

Since most of the actual details of the performance-based assessment remain ideas and have yet to be developed, future directions for pupil services professionals are not clear at this time. However, the following recommendations are meant to serve as a list of practical suggestions for pupil service professionals in terms of preparing for the eventual implementation of performance-based assessment in their school systems:

1. Begin by studying the Kentucky Education Reform Act, Part I-Curriculum, specifically the sections that define and discuss performance-based assessment.
2. Become familiar with performance-based assessment as an alternative to standardized, pencil and paper, multiple-choice format tests. Consult with university and college faculty with expertise in testing and measurement concepts. Identify resource materials on this topic. Begin a school resource file with information that can be made available or distributed to school staff.
3. Begin a dialogue with school administrators and colleagues about performance-based assessment. Attempt to clarify what roles and responsibilities pupil service professionals might assume with the implementation of the proposed assessment plan.
4. Assist your schools in identifying the affective as well as cognitive outcomes that will be evaluated through performance-based assessment as specified in K.E.R.A. Encourage teachers to address these goals and outcomes through their classroom instruction.
5. Offer instructional consultation to classroom teachers on ways to incorporate curriculum-based performance evaluations into their classrooms.
6. Conduct workshops to inform teachers, parents, and students about performance-based assessment. Clarify misunderstandings. Emphasize the value of meaningful assessment. And reduce stress and anxiety due to uncertainty and misinformation. Focus on the reward aspects of the reform law rather than the threat of sanctions.
7. Establish channels of communication with county and state administrators, as well as legislators, in order to provide feedback and offer suggestions concerning the development and implementation of performance-based assessment.

Mathematics

William S. Bush

Sheila Vice

The morning session began with an attempt by the group to define performance-based assessment. In their definitions, members of the group used words and phrases such as: hands-on, active, practical, open-ended, subjective, the process of getting right answers vs. just the right answers, using materials, and not "multiple choice." The discussion then focused on the term "authentic." The group raised the following questions: Does "authentic" involve only consumer-related tasks relevant to all citizens in their daily lives? Does "authentic" mean the same for average citizens, college-bound students, or mathematicians? The group decided that mathematics tasks exist which are "authentic" for college-bound students and mathematicians, but cannot be placed in a consumer-oriented situation. In essence, the group suggested that many mathematics activities are important for their own sake and for developing thinking skills.

After this discussion, several concerns regarding performance-based assessment were raised by group members. The first concern was the emphasis on reading in mathematics tasks. Some group members were concerned that performance assessments would measure reading comprehension rather than mathematical expertise. They felt that poor readers would also do poorly on mathematics tasks requiring extensive reading.

A second concern was that the length of the tasks would cause students to become tired and perform poorly. Some members felt that students could not concentrate for extended periods of time.

A third concern was that tasks would be biased and prejudicial. Members were concerned that once tasks are placed in a specific context, then those students with out-of-school experiences in those contexts would have an advantage over students with limited experiences. They felt that performance would be affected by SES and parental involvement with their children, rather than what teachers taught in the classroom. The group members were shocked and dismayed by Gary Phillip's comments that minority students tended to score more poorly as the tasks became more "authentic."

A forth concern centered on the random sampling of students to complete the performance-based assessment. Group members felt that test scores in a district or school could be affected by the particular students who are randomly selected to take the assessments. Some group members wondered what students who were not being assessed would do while the randomly selected students were taking the tests. The group felt that all students in a school or district should be assessed.

A fifth concern was that only one type of assessment would be used. Some students perform better on multiple choice tests; others perform better on open-ended tests. The group felt that a variety of assessments--multiple choice, performance based, projects, interviews, etc.--be used to assess mathematics learning. Some group members felt that basic skills should be assessed first and separately from performance-based tasks.

The morning session closed with the two group leaders sharing several examples of performance based assessment. Each group member received a copy of a sample mathematics performance task from (1) the Kentucky Mathematics Task Force of the Council on School Performance Standards, (2) the California Statewide

Assessment, and (3) the Connecticut State Assessment. Dr. Bush outlined the assessment procedures for college entrance in the Victoria (Australia) Ministry of Education. This assessment consists of four parts (projects, problem solving, mechanics, and concepts) given at four different times during students' final year of pre-college schooling.

Because of the light attendance, this session was more conversational and covered many different issues. As in the morning session, the vast majority of the discussion focused on specific concerns regarding the upcoming assessment in Kentucky. The first concern was that 12th grade students will not take the tests seriously, and, in some cases, might try to do poorly on the tests. The group felt that eleventh-grade students would be more motivated to perform well on the tests.

BEST COPY AVAILABLE

Educators in the afternoon session were concerned about the length of the assessment. Most members felt that day-long and three-day tasks were inappropriate for students. They were concerned that this type of assessment would be inefficient, expensive, and very time-consuming. The group also voiced the same concerns as the morning group regarding random selection of students.

Group members were also concerned with scoring. They felt the scoring on such tasks would not be objective or fair. They were also concerned with the plan to use baseline data and compare subsequent scores with the baseline. Several group members felt that this format would encourage districts to "cheat," and students to perform poorly, on the initial administration of the tasks.

The group also felt that the Kentucky plan for performance-based assessment would be very expensive. The development and scoring of tasks and the hiring of assessors and scorers will cost the taxpayers of Kentucky a great deal of money.

Finally, the group felt that teachers will require retraining in both teaching and assessment techniques. Teachers will need new and different types of teaching materials. In addition, administrators, parents, and students will need training and awareness workshops.

In summary, the school administrators and teachers attending this session were scared. They are afraid that performance-based assessment will force them to do things for which they have not been prepared. They are afraid that they will be held accountable for results and performance for which they have no control. They are afraid that they will not have not have the resources and training to do what is necessary to help students perform well on performance-based assessment. On the other hand, many educators were excited about new techniques and assessment which will make learning and teaching fun and enjoyable. If they just were not being held accountable.....

Foreign Languages

Phil L. Nacke

Linda Kraus Worley

What is performance assessment?

The participants in the section on foreign languages agreed:

- that all four skills (reading, writing, speaking, and listening) as well as culture need to be assessed,
- that the tasks need to be realistic and very well defined,
- that students need to be assessed as to how well they can function in real-life contexts,
- that the assessment procedure needs to look at how well grammar principles are applied in these real-life situations and that students need to demonstrate more than merely a knowledge of grammar rules, conjugations, declensions, etc.,
- that the assessment is keyed into levels of proficiency as defined by the American Council on the Teaching of Languages,
- that the scoring system needs to take into account both the complexity of the task and use, at least in part, holistic scoring criteria,
- that the assessment procedure should integrate lower levels of knowledge into higher levels, recognizing that language proficiency is to be rated on a continuuum.

Given illustrations of performance assessment in this area, what are the criteria for deciding whether they are good or bad?

While keeping in mind the definition of performance assessment outlined above, the group focused on the following issues:

If the performance assessment is to be based on real-life situations, then there must be oral/aural and reading/writing components which are tied in with ongoing classroom concerns and experiences. Such communicative experiences will have a cumulative effect on student proficiency.

There is the recognition that a sampling of performance will give an efficient and valid measure of student proficiency.

A performance-based assessment should have the effect of providing valuable feedback to the classroom teacher and the learners as to the efficacy of curriculum and classroom practice in light of real-world expectations of language fluency. If these real-world expectations form the basis of assessment, then these assessment procedures will also satisfy the need for accountability evidenced on both the district and state-wide levels.

A good performance-based test will motivate students to do well in that it will present students with interesting and meaningful tasks to be accomplished.

Some characteristics of valid and efficient tests are that the tests will:

- be contextualized
- be holistic
- be criterion-referenced
- be able to be successfully completed in the time available
- be comprehensive and flexible enough to allow for the diversity of vocabulary students will have learned
- use authentic language
- give students the sense that the tasks have benefitted them
- give students the sense that the tasks really test what they know, i.e., that the test is an appropriate, adequate measure of learning
- give students the chance to demonstrate what they can do

The group recognized that, although the speaking component of any performance-based test may be more difficult and labor intensive to test than the other skills, there are methods available to test speaking in an efficient manner.

Given examples of current practice in this area, in what sense have they worked well?

The Guidelines of the American Council on the Teaching of Languages were cited as a proficiency scale underlying much current performance-based testing. (ACTFL Proficiency Guildelines. Hastings-on-Hudson, NY: ACTFL Materials Center, 1986.)

Are there implications of performance assessment in terms of fairly assessing the different genders, racial and ethnic groups, and social classes?

The group discussed the issue of cultural diversity and performance outcomes in terms of:

- lexica, concrete topics to be tested, etc.
- the fact that a foreign language can serve to motivate students due to the fact that students tend to begin with zero knowledge of the language. This fact can place students from diverse backgrounds on a more equal level.
- by teaching the culture of the foreign language, students should become more sensitized to cultural diversity.

What are the structural barriers to using these activities in the schools? Are there areas in which training will be required?

The group felt that there were no structural barriers per se to implementing these new kinds of testing procedures, but felt that:

- teachers would need to be retrained in terms of the communicative methodologies which are performance-oriented.
- teachers will need to have access to the materials necessary to focus students on real-life situations.
- teachers will also need to be trained in methods that will allow learners to use the language in real-life groupings, i.e. with one or more conversational partners and that will optimally involve each student in using the target language.
- teachers may also need to have extra training to increase their own foreign-language skills, especially their oral skills.
- the Commonwealth of Kentucky needs in-service standards reflecting the skill levels necessary for teachers in a proficiency-based, communicative classroom.
- the Commonwealth of Kentucky must provide the financing necessary for workshops and other training needed to bring teachers up to the necessary standards.

Instruction and Assessment

Thomas Guskey

We now turn our attention to the initial tasks involved in planning for the implementation of mastery learning. The first of these tasks is to set out, in specific terms, what students are expected to learn. These expectations are generally referred to as *learning objectives*, and they are the main topic of this chapter.

Learning objectives describe the skills and abilities students are to acquire as a result of our teaching. The process of specifying them requires that a series of important decisions be make about what is essential for students to learn and at what level that learning should occur. While these decisions are fundamental to the use of mastery learning, they also add precision to procedures for evaluating students' learning. In mastery learning they further serve as a basis for developing formative tests and planning feedback and corrective activities.

To assist teachers in making these decisions, many commercial publishers list the particular learning objectives that their materials have been designed to help students attain. Although these vary widely in detail, most teachers find them helpful in planning instruction. Some commercial materials also include tests for checking on students' learning progress. These, too, can be useful and can help reduce the amount of preparation time required of teachers. Unfortunately, though, few commercial materials are universally applicable. Refinements and adaptations are usually necessary to meet the needs of particular groups of students or the instructional preferences of different teachers. Therefore, along with considering the decisions that need to be make in outlining learning objectives, we will in this chapter also consider procedures for reviewing and adapting objectives and materials that may be already available.

THE IMPORTANCE OF A STRUCTURE FOR LEARNING

Generally when scholars or researchers reach a certain level of sophistication in a subject, they are able to see definite relations among the ideas and concepts in that subject. These relations help them understand more complex phenomena and aid them in conducting further study. In many cases, curriculum writers, most of whom are also experts in the subject, make the assumption that the *structure of the knowledge* in that subject is synonymous with an appropriate *structure for teaching and learning* that subject. Unfortunately this is not always true.

For example, several modern mathematics and science curricula were based upon the belief that if the sophisticated organizing principles that experts found so helpful could be provided to young people learning the subject, the young people would find the subject much easier to learn. It was quickly discovered, however, that while these organizing principles are very useful for specialized scholarship, they are not always useful in helping the majority of students learn the subject. As B.S. Bloom, J.T. Hastings, and G.F. Madaus point out:

> The usefulness of a structure for learning has to do with the ability of students to comprehend it and use it as an organizing factor in their learning. There is no relation-between the usefulness of a structure for scholars and its usefulness (and meaningfulness) for students. (1971. p.12).

It is true that students learn more easily when provided with a structure that helps them relate various aspects of the subject. Such a structure also helps students gain deeper meaning from what might otherwise be a large number of unconnected specifics. Research studies, particularly those on the use of "advanced organizers" in instruction (Ausubel, 1963, 1978), have shown that ideas are more easily grasped and remembered when learned in relation to one another rather than in isolation.* A structure for learning should thus provide students with a mechanism they can use to better understand the instruction and to organize the concepts they are learning. It is important to keep in mind, however, that a structure for learning is based primarily on pedagogical considerations and may not be the same as an expert's or scholar's view of the field.

ORGANIZING LEARNING UNITS

Developing an appropriate structure for learning generally involves three elements.

1. *The final learning goal to be attained must be specified.* This goal is usually a competent learner who has truly mastered those things we set out to teach.
2. *The final learning goal must be analyzed to identify the steps that are necessary to reach the goal.*
3. *The steps must be ordered in an appropriate sequence to facilitate learning and provide for steady and regular progress toward the goal.*

Although these three elements may seem implicit in all teaching, one or more is often neglected. For instance, the daily burdens of teaching can sometimes distract a teacher's perspective from the final learning goal and, as a result, teaching efforts lose their focus and cohesion. Similarly, concentrating solely on the goal without careful attention to the separate steps required to reach that goal can result infrustration for teachers and students alike. Both the goal and the sequence of steps required to reach that goal need to be kept in mind if teaching and learning are to be effective.

The process of analyzing a learning goal and then organizing the steps necessary to reach that goal is a natural part of most teaching and learning activities. For example, consider how you might go about teaching a child to play tennis. You would probably begin with a mental picture of the child approaching the ball, swinging smoothly, and returning the ball to the other court. This is the final goal you would hope to attain at the end of the learning process. From that mental picture you would begin by dividing the components of that final learning goal into various steps. You would probably think about adjusting the racket to the child's size and strength; adjusting the child's grip for backhand and forehand returns; telling the child about the importance of watching the ball; and showing the child how to move to the ball in order to make the backswing, return and follow-through, and then recover for the next return. You would also need to demonstrate serving. In addition, you would want to explain the rules of the game and how to keep score. Given this breakdown you would then decide upon an appropriate sequence of learning steps. You might decide to order the steps in terms of difficulty or complexity. The most basic elements, such as watching the ball, would be presented before more complicated steps, such as an appropriate follow-through. Then, as you begin teaching, you would try to be aware of any special problems the child may be having and would try to correct them as they appear. In addition, you would probably make a point of rewarding the child whenever possible and providing reassurances at other times.

This example illustrates the process that takes place in most effective teaching and learning situations. The learning goal, or what is sometimes referred to as the *summative goal*, is first analyzed in terms of the parts that need to be mastered. Those parts are then organized and arranged in an appropriate sequence of learning steps. Care is taken to ensure that each of these steps is mastered while progressing toward that final goal.

Similarly, to begin the use of mastery learning, it is first necessary to identify the final learning goal and to define the specific steps that need to be mastered in order to reach that goal. Many teachers do this regularly as a part of their instructional planning. That is, they start out with a mental picture of a competent learner at the end of the course of instruction and then divide the material that must be learned over the year or term to reach that goal into smaller components or steps. Each of these steps is considered a *learning unit.*

The delineation of learning units is somewhat arbitrary in many cases. Ideally, learning units should be determined by natural breaks in the subject material or by content elements that make a meaningful whole. Thus each unit might not cover exactly the same amount of content. Textbook publishers usually divide the content of a particular subject in accordance with these natural breaks. For example, chapters in textbooks often represent appropriate learning units.

Another critical element to keep in mind in determining learning units is instructional time. A learning unit should contain material that can be presented in about a week or two of classroom time. Generally, learning units at the high school or college level are longer and cover more material than learning units in the elementary grades. A unit in a high school course may last two weeks or slightly longer and cover twenty or thirty important objectives. An early elementary school unit, on the other hand, seldom lasts more than a week and may cover a single skill or fragmented and generalizations will be difficult to build. If too long, however, students who fall behind because of particular learn-

ing problems may have great difficulty catching up. Thus not only the content but also the pace of instruction and the kinds of students involved in the learning need to be considered in determining appropriate learning units.

TABLES OF SPECIFICATIONS

Once the sequence of learning units is clearly delineated, the next task is to specify the learning objectives of each unit. To do this, it is necessary to first identify the new material that will be presented in each unit. Many teachers develop detailed outlines of the new material they plan to present and often consider this material to be the objective of the unit. But while detailed outlines of new material can be very useful in teaching, they say nothing about what students might be expected to do with that material. Therefore, in specifying learning objectives for a unit, it is important to consider not only the material or *content* students are expected to learn, but also the specific *behaviors* they are to attain in relation to that content. These behaviors indicate the ways we would like students to be able to think, act, or feel about the material, about themselves, about others, and so forth. That is, what students are expected to do with the material is just as important a consideration as the material itself.

A useful and efficient way to outline the new material and the behaviors students are expected to attain is to construct a simple two-dimensional table. This table is sometimes called a "mastery chart" or a "matrix of content and behaviors." More generally, it referred to as a *table of specifications*.

A table of specifications is basically an outline of the learning objectives for a unit. As such it serves two important functions. First, it helps add precision and clarity to teaching. The information on the table should be precise enough to convey exactly what is intended in the instruction, and clear enough that students can be helped to fully understand what they are expected to learn. Many teachers go through a similar specification process as a regular part of their class preparations. But often what is expected of students is never make clear until the time they are evaluated. In these instances, students are forced to guess what is important and what is inconsequential. While some students are very good at guessing what they are expected to learn, many others are not.

For those who are not, learning soon becomes a very frustrating experience. Specifying the objectives for learning precisely and clearly, and communicating these objectives to students, not only eliminates much of this guesswork but also helps organize and focus teaching and learning activities.

The second function of a table of specifications is to serve as a guide for consistency between learning objectives and procedures for checking on students' learning progress. Although this kind of consistency is very important for learning, it is often neglected or given only cursory attention. For example, many classroom teachers stress that they want their students to develop higher cognitive skills such as the ability to make applications, analyses, or syntheses. However, the vast majority of classroom tests tap only those skills that are easiest to assess, such as knowledge of the definitions of terms or specific facts. A table of specifications can be used as a guide in preparing a wide variety of tests and evaluation procedures. In this way, it can help to guarantee consistency between important objectives and procedures for checking on students' learning (see, for example, the sample tables in chapters 9 through 14).

STEPS IN DEVELOPING A TABLE OF SPECIFICATIONS

The first step in preparing a table of specifications is to determine what new material or content is introduced in the learning unit. That is, what are the new terms, facts, relations, procedures, and so forth that are explained, defined, illustrated, or otherwise presented in the unit? Usually textbooks and other learning resources are relatively clear in signaling when new material is being introduced. Changes in print or color, comments in the margins (particularly in teachers' guides), and summaries at the end of chapters often identify new content.

The next step is to determine the particular student behaviors that should be paired with the new material or content. That is, what are students expected to do with the new content? Will students be required to simply remember or recall the new content, or will we want them to be able to use it in a new or different way? In specifying these decisions, many teachers find it useful to classify the new elements of content according to some of the categories in *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain* (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956). These

categories represent a hierarchy of levels that differ in terms if difficulty and complexity. The lowest levels represent the simplest kinds of learning, while higher levels represent more advanced cognitive skills. The categories that are most useful in a wide variety of subject areas are shown in the table of specifications in Figure 2.1. These levels include:

1. *Knowledge of terms.* Terms are the new words or phrases that students are expected to learn. They may be expected to define these terms, recognize illustrations of them, determine when they are used correctly, and/or recognize synonyms. Examples of such terms are *product* in a mathematics unit on multiplication or *photosynthesis* in a science unit. The "knowledge" level here is simple recognition or recall, and is generally considered to be the simplest level of student learning. All new words or phrases to be introduced and explained in the unit are listed in the table under "Terms."

2. *Knowledge of facts.* Facts are the specific types of information that students are expected to remember. In general, facts are particular details that are important in their own right or are essential for other kinds of learning. Facts include names of persons, events, operations, or other kinds of specific information. Students may be expected to recall particular facts and/or remember the correct fact when asked about it in a relatively direct manner. Examples of facts are "The Declaration of Independence was signed on July 4, 1776" and "Stephen Crane is the author of *The Red Badge of Courage.*" Any new facts in the unit should be listed in the column labeled "Facts."

3. *Knowledge of rules and principles.* Rules and principles concern specific patterns or schemes that are used to organize the major ideas of a subject. Generally they bring together a number of facts or describe the interrelationships among a number of specifics. Students may be expected to know a rule or principle, to remember an illustration of it used during instruction, and/or to recall situations in which it was applied. In most cases, rules and principles are more difficult to learn than are terms or facts. However, it is important keep in mind that this category deals only with the knowledge of rules and principles, not with their application. The "commutative principle" in mathematics and the "rules for subject/verb agreement" in grammar are examples of rules and principles. New rules and principles in a unit should be listed in the table.

4. *Knowledge of processes and procedures.* In many subjects, students are expected to know the particular steps involved in a certain process or procedure. Frequently it is important that these steps be recalled in a specific sequence. For example, students may be expected to know the appropriate order of steps in a mathematics problems-solving task and/or the sequence of events necessary to enact legislation in Congress. Processes and procedures may involve a number of terms or facts and are typically difficult for students to learn. Any new processes or procedures should be included in the table. It is important to keep in mind, however, that these should be the processes and procedures we want students to learn, *not* the instructional processes or procedures we plan to use in teaching the unit.

5. *Ability to make translations.* Translation involves the transformation of a term, fact, rule, or process from one form to another. In making a translation, students express particular ideas in a new way or take phenomena or events in one form and represent them in an equivalent form. Students

TABLE OF SPECIFICATIONS						
Knowledge of						
Terms	Facts	Rules and Principles	Processes and Procedures	Translations	Applications	Analyses and Syntheses

may be expected to recognize new illustrations of a term, fact, rule, or other matter and/or to determine whether a new illustrations is appropriate. In general, then, students employ translation when they put an idea in their own words or recognize new examples of what they have already learned. Translation in this sense should not be confused with translation in foreign language instruction, though they are similar. (The special considerations regarding foreign language instruction are discussed in detail in Chapter 14.) The specific translation skills students will be expected to learn should also be listed in the table.

6. *Ability to make applications.* Application is the use of terms, facts, principles, or procedures to solve problems in new or unfamiliar situations. Students are expected to use ideas or concepts learned in one context to solve a problem presented in a new context. It is important to remember, however, that if the problem is one students have encountered previously except that new data are substituted, then the behavior called for is really a translation rather than an application. Because of this , it is difficult to name specific examples. An application for one teacher might be a translation for another, depending upon differences in the examples used in class presentations or practices. In making an application, students must first recognize the essentials for the new problem; determine the facts, rules, procedures, and so forth that are relevant; and then use these to solve the problem. The ability to make applications is a fairly complex behavior and is the highest level of learning objective used by most teachers. The application skills students are expected to develop in a unit should be included in the table.

7. *Skill in making analyses and syntheses.* An even more advanced behavior is that of making analyses and syntheses. Analyses generally involve the breakdown of concepts into their constituent parts and the detection of the relationships among those parts. Syntheses, on the other hand, involve the putting together of elements or concepts in such a way as to develop a meaningful pattern or structure. Syntheses often call for students to develop creative solutions within the limits of a particular problem or methodological framework. Because of the complexity of these behaviors, analyses and syntheses are typically considered only in advanced or higher level classes. However, some teachers feel that these skills are very important and try to include practice in these kinds of activities for students at all levels. Examples of analyses are "distinguishing facts from opinions in a communication" and "identifying conclusions and supporting statements." An example of synthesis is "Write a paragraph that organizes a set of ideas and statements." Any analyses and syntheses included as a part of students' learning should be described in the table.

In general, teachers have little difficulty outlining learning objectives in terms of these categories of behavior. Most find that new material in a unit can usually be classified in one or another of the first four knowledge level categories. Few textbooks or other learning resources present material at more advanced levels. Decisions regarding more advanced behaviors must therefore be made apart from the materials.

At times, some confusion does arise as to whether a particular objective should be listed under one or another of two adjacent categories. For example, an objective might be considered a fact by one teacher and a principle by another; or a translation to one teacher might be considered an application by another. But differences such as these are usually of little consequence. So long as the objective is included in the table, whether it is in one column or the next adjacent column makes little difference. Major distinctions, for instance, between the knowledge level and the application level, are important, however, and need to be carefully considered.

RELATIONSHIPS AMONG LEARNING OBJECTIVES

Besides listing the new elements of content and what students are expected to do with that content, the table of specifications can also be used to illustrate relationships among content elements. For instance, knowing the definition of a term may be necessary in order to understand a fact pertaining to that term. Or, knowing two or three facts may be essential in understanding a particular procedure. These relationships can be illustrated on the table by drawing connecting lines between these elements. For example, lines might be drawn from two or three new terms to a fact that incorporates these terms. A line could also be drawn from knowledge of a particular procedure to the application of that procedure. Many teachers find that drawing these lines helps them keep these relationships in mind so that they can be developed for students. An example

illustrating these relationships for an elementary social studies unit is shown in Figure 2.2.

For the vast majority of teachers, developing a table of specifications is not only useful, it is also very revealing. The table enables them to view the new elements and their relationships in a very compact form. By making tables of their own or by carefully reviewing prepared tables, teachers can add greater precision to their teaching and can more closely match presentations of unit material with learning objectives. A table of specifications can also help reveal possible gaps in the instruction. It can show where important elements may have been neglected or where relationships between elements need to be pointed out. Furthermore, a table is useful in constructing instruments to check on students' learning progress since it provides a guide to what elements of content should be tested and what student behaviors should be assessed. These considerations are discussed more extensively in the next chapter.

TABLE OF SPECIFICATIONS (Partial)

	Knowledge of	
Terms	Facts	Rules and Principles
Geography – Geographer Map Scale Longitude Latitude Eratoshenes map Ptolemy's map	Where you live affects the things you do The skill of mapmaking is very old. People have made maps based on in- - accurate information Inaccurate maps may affect explanations.	Earth features influence many human activities: -the location of towns and cities. -the routes traveled. -the things eaten. -occupantions.

Figure 2.2 Portion of table of specifications showing related elements of content in an elementary social studies unit.

SUBJECT AREA DIFFERENCES

Learning objectives from nearly any subject can be outlined according to the categories shown on the table of specifications in Figure 2.1. However, all of these categories may not be applicable to a particular subject. Some subjects, such as social studies, may have no or very few rules and principles. Similarly, some units within a subject, such as mathematics, may not include any new terms. Thus it is not unusual to find that some of the categories in a table of specifications for a particular subject or unit are omitted or left blank. Only those categories that are useful in listing the new material and specifying the behaviors expected of students need to be included. These subject differences can be seen by inspecting the tables of specifications in Chapters 9 through 14.

The difficulty of preparing a table of specifications also varies depending upon the subject. Some subjects, such as mathematics and science, lend themselves quite easily to this format. It is usually not difficult to identify the new terms, facts, rules, and/or procedures that are part of a mathematics or science unit. However, subjects such as reading, composition, and language arts are not as easily categorized. Special adaptations are sometimes necessary for these subjects, as can be seen in the language arts unit in Chapter 11. Still, it remains important to precisely specify the new ideas or concepts to be presented, and what students will be expected to do with that information. For instance, in teaching students how to make inferences, it is

important to specify the various components that go into making inferences and how to distinguish between a student who can make appropriate inferences and one who cannot. And more importantly, we must consider how to help students who cannot make inferences to acquire that skill. Issues such as these need to be given serious consideration when outlining learning objectives for a unit in reading or composition. This kind of specification greatly enhances the organization of instruction and helps students be more focused in their learning efforts.

TABLES OF SPECIFICATIONS AND THE TEXTBOOK

For most teachers, the basic textbook used in teaching serves as a primary guide to the new material introduced in a learning unit. Some teachers do supplement the material presented in the textbook with information from curriculum guides, other textbooks, learning kits, or special classroom activities. But, generally, the course textbook is the principal source of new material presented in each unit.

The vast majority of textbooks present new material at strictly the knowledge level. In fact, few textbooks compel students to do anything other than know or recall the information presented. Occasionally at the end of a chapter there will be a section entitled "Questions for Discussion" or "Special Activities" that requires students to use the material presented in the chapter in a new or different way. But since so few textbooks go beyond the knowledge level when presenting new information, often only the knowledge-level categories are filled when a textbook is used as the principal guide in initially preparing a table of specifications.

Nevertheless, most teachers are interested in a wider range of student behaviors than simply those associated with basic knowledge. Many want their students to develop higher cognitive behaviors as well, to use what they are learning in new and creative ways. In these instances, it is necessary to go beyond the textbook in outlining learning objectives. That is, although the textbook can serve as an excellent guide to the content of a learning unit, it usually does not clearly specify what students might be expected to do with that content. These decisions must be made by curriculum planners or by individual teachers. In preparing tables of specifications or reviewing tables that have been previously developed teachers need to consider the types of behaviors they want *their* students to attain as a result of *their* teaching.

TABLES OF SPECIFICATIONS AND LESSON PLANS

The process of preparing a table of specifications can greatly add to the clarity of lessons and class presentations. In fact, many teachers use the table as a guide during teaching activities. But for many others, a table of specifications by itself is insufficient as a lesson plan.

Particularly at the elementary level, most teachers prefer to teach from lesson plans that are fairly detailed. These plans describe not only *what* is going to be taught but also *how* it will be taught. In preparing these plans, teachers often list the pivotal questions they plan to employ in their presentations or the guiding questions they want to use to stimulate discussion. Sample problems may be illustrated and specific examples may be listed. While details such as these are extremely useful in presenting a lesson, they are usually not included in a table of specifications. A table of specifications is usually quite brief and will describe the learning objectives for an entire unit in a page or two. In essence, it is designed to address two questions: (1) What do I want my students to learn? and (2) What do I expect my students to be able to do with what they have learned? The first of these questions concerns the content; the second concerns student behaviors in relation to that content. Note, however, that both of these are "what" questions. The "how" questions are usually not addressed in the table.

It is true that the types of learning outcomes described by the objectives outlined in a table of specifications suggest different approaches to teaching. For example, while drill and recitation may be appropriate techniques for teaching knowledge of terms and facts, they are inappropriate for teaching students the skills involved in making analyses and syntheses (Gagne, 1977). The inclusion of both of these levels of student behavior on a table of specifications thus implies that different approaches to instruction will be used. But, generally, teachers who want a more detailed plan that illustrates specifications with lesson plans that include further directions and information. Such lesson plans typically describe the activities to be used each day or in each class session, and they directly address the question of "how" these objectives are to be accomplished.

BENEFITS OF A TABLE OF SPECIFICATIONS

Outlining learning objectives and preparing tables of specifications clearly illustrate the importance of the decisions make by teachers in using mastery learning. Even the process of reviewing tables prepared by other teachers or curriculum specialists requires that careful thought and consideration be given to what students will be expected to learn. Although the importance of these decisions are a fundamental part of all teaching. Judgements about what should be taught, what students should learn, and what students should be expected to do with what they have learned are made by teachers on a daily basis. Developing a table of specifications simply compels teachers to be very conscious of those judgements and decisions, and to make them in a very explicit way. As a result teaching activities can be better organized and much more focused.

The process of preparing a table of specifications may seem somewhat difficult and cumbersome at first. Generally this is because making these judgements and decisions in such a specific manner is an unfamiliar experience. The process requires a perspective toward learning objectives and instructional goals that is new for many teachers. However, within a very short time most find that the process becomes much easier. The time required to prepare a table of specifications for a second unit is usually about half that required for the first. A table for a third unit requires even less time.

In addition, most teachers discover that teaching becomes somewhat easier as a result of this process. In the long run, clearly specifying what students are to learn and what they are expected to do can save a teacher valuable time and also provide a sense of accountability. Furthermore, being well organized allows a teacher to concentrate more fully on ways to best present new material and on different methods for involving students in the instruction. In this way, a table of specifications is similar to a map used by travelers. While it does not limit the pathways that can be taken, it does enhance the efficiency of the traveling, the enjoyment of the travelers, and the likelihood that all will successfully reach a particular destination.

SUMMARY

Having a clear notion of what is to be taught and what students are expected to learn is essential in implementing mastery learning. This is usually accomplished by outlining specific learning objectives. The tasks involved in this process are:

1. *Identify specific learning units.* Learning units are typically determined by the natural breaks in the subject content and usually represent the material covered in about a week or two of instructional time.
2. *Determine the content to be presented in each unit.* List the new terms to be introduced, the facts to be presented, the rules and principles to be explained, and the processes and procedures to be discussed.
3. *Determine what students are to do with the new content in each learning unit.* Indicate whether students will be expected to know or recall the new material, translate it from one form to another, apply it in a new or different situation, or analyze it in a new way.
4. *Prepare a table of specifications summarizing these decisions.* A table of specifications outlines the content of a unit and the learning outcomes expected of students in relation to that content. It can also serve as a guide in planning instruction and evaluating students' learning.

ACTIVITIES

1. In a familiar subject area, preferably one you teach, identify a learning unit. If you do not presently teach, consider the material presented in this chapter as a learning unit.
2. Construct, using the general format shown in Figure 2.1, a table of specifications for the unit.
3. Have a friend or teaching colleague review, both the unit and your table of specifications, asking of the table, "Are these the same elements of content I would identify a important in this unit"? and "Are these the same behaviors I would require of students?"
4. Discuss any differences that arise through this review. Although such differences are likely to be few, their clarification can add greatly to the precision and completeness of a table of specifications.

REFERENCES

Ausubel, D.P. (1963). *The psychology of meaningful verbal learning.* New York: Grune & Stratton.

Ausebel, D.P. (1978). In defense of advance organizers: A reply to the critic. *Review of Educational Research*, **48**, 251-257.

Bloom, B.S., Englehart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of educational objectives, handbook I: Cognitive domain,* New York: McKay.

Social Studies

Linda Levstik

Sharon Brennan

Fifteen attended the morning session including the two leaders. There seemed to be some distribution geographically and professionally with a mix of administrators and teachers from various school districts in the eastern half of the state.

Linda Levstik opened the session with a round of introductions and a reaction to the remarks made by Dr. Gary Phillips about the NAEP program. Linda explained that the NAEP Social Studies test is not performance based. She then showed samples of the performance assessment items which are under consideration by the state Social Studies Task Force for grades 4 and 8. Discussion center on reactions to the items shown and the following questions:

What is performance assessment?

This question raised concern among participants about potential problems with performance based assessment. One participant mentioned that while it is laudable that performance based tests aim to be authentic, it is important to remember that many activities in the classroom cannot be assessed. Others expressed concern about how to design tests that capture students' knowledge of and ability to interpret content and simultaneously test process skills. They also wondered how schools might be compared under this system.

Other potential problems mentioned include the impact of mainstreaming on curricular changes and the difficulty involved in creating objective, valid, reliable tests. One participant expressed concern about administration of the test. If these tests are given over several days, validity could be jeopardized since students would have opportunities to seek assistance. Another participant felt it would be difficult to test both content and process skills with this measurement system.

One participant suggested that policy-making bodies should focus on examining the knowledge base which undergirds the Social Studies curriculum. This person also suggested that the linkage between pedagogy and content should be considered as well as how to integrate Social Studies with other areas of curriculum

What are the criteria for assessing assessment?

Participants agreed that the most important criterion is to devise a way to test process skills and content together. They also thought a careful determination of which content areas should be tested is a critical consideration since Social Studies encompasses such a broad area of topics and issues. Several felt that specific goals should be determined and a variety of tests given in order to assess all aspects and levels of learning from retention to interpretation of subject matter.

How does performance assessment compare with traditional "paper and pencil" tests in terms of validity or efficiency?

Participants agreed that multiple choice tests are more efficient. The utility of performance assessment tests is yet to be determined. Participants felt they would need to have more information in order to respond adequately to this item, and concern was expressed about the lack of information which has been provided, to date, about implementation.

Given examples of current practice in this area, in what sense have they worked well?

Participants felt that, at this point, they do not have enough experience to determine how performance assessment will work; however, the following comments were made about the issue in general.

One participant pointed out that many reform initiatives can be traced to John Dewey's work and the Progressive movement in education in the early 1900's. Another participant reiterated the importance of establishing meaningful goals and adapting the curriculum to fit these goals. Someone else suggested that in doing so careful consideration should be given to the selection of material used for instruction, especially textbooks. Another par-

ticipant suggested that those designing assessment measures need to work closely with the State Textbook Commission, because the Commission, through the selection of classroom material, sets the curriculum. She also said that it will be important to acquaint the public with changes in assessment and the resulting change in pedagogy.

What are the implications for fairly assessing different gender, racial, ethnic groups and/or social classes?

Discussion of this issue centered around Dr. Phillip's comment that minority students do not score well on performance assessment type items.

The point was made that students from varied ethnic and social groups tend to score poorly on the Social Studies performance assessment items, but black and hispanic students score very poorly. Participants speculated that the reason for this may be lack of experience with reading and writing. One participant suggested that tests should account for differences in learning styles.

Seven participants attended the afternoon session including the two leaders. As with the morning session, this began with a round of introductions and orientation by sharing two samples from the state task force's work. The discussion centered on reacting to these items and responding to the following questions:

Do the activities defined, illustrated, and discussed in the morning session meet the requirements of the law?

Generally, participants felt that teachers are engaged in the type of activities that fit with performance assessment although some noted that not all teachers create activities that would fit. Participants suggested that, in order to create a fit, extensive training would be needed.

What are the structural barriers, if any, to using these activities in the schools?

One barrier is that Social Studies is often left out of the curriculum, and the Social Studies curriculum for the primary grades tends to be weak. A second barrier, according to participants, is parent involvement. Parents want justification for changes in testing and curricula, and if they are not satisfied with the reasons for change, they may voice objections. Departmentalization was the third barrier mentioned. Schools need to integrate social studies into other aspects of the curriculum and vice versa. Textbooks are also problematic because teachers tend to use them as a crutch. Teachers need to consider using a wide range of material and activities to interest students in learning about Social Studies. A fourth, and final, barrier is the amount of content to be covered in the Social Studies curriculum. Participants felt that content would have to be carefully selected.

What are the costs of implementing these new activities in terms of both material and labor?

Implementation will be costly in terms of both time and labor, according to the participants. Social Studies requires use of a wide range of material, and the amount of content requires a great deal of time to cover. In addition, longitudinal studies will be needed to determine the impact of performance assessment as students progress through the grades. One participant pointed out that it would take half of an average teaching career to study one group of students grades K-12.

Are there areas in which training will be required?

Extensive training is needed for implementation of the reform initiative in the area of Social Studies. Participants felt that teachers need to have time away from their school duties to use performance type tasks and discuss practices which would help students learn. One participant suggested that a very effective form of training would be to have consultants teach a class, demonstrating methods over an extended period of time.

Summary

Emanuel Mason

From the discussions in the various sessions and presentations given at the conference, it appears that professionals in the schools still do not know very much about how performance assessment will be done in Kentucky. Yet we only have a short time to prepare for performance assessment in the Kentucky schools. The lack of direct and simple implementation suggest that part of the implementation include funding for training for teachers and school administrators.

Based on the experience in this and other states and in NAEP, certain useful points can be made as we prepare for performance assessment in our schools. KERA mandates changes in school assessment from an emphasis on relatively artificial demonstrations of knowledge (like paper-and-pencil multiple-choice tests) to the seemingly more practical performance based assessment. Yet assessing school performance using so-called realistic and practical tasks is not as obvious or as simple as it appears on first glance.

Further, what this means for teaching and the classroom is different from one field to another. For example, physical education, vocational educational, language education, and much of art education have been largely peformance-based for some time. Other fields like math, social studies, and science will have to make more significant changes in how they view assessment. So the impact should be different in different areas of the curriculum.

Tasks in a performance assessment should be varied. Some might be very practical, like balancing a checkbook or creating a recipe for a blueberry cobbler. Others might be paper-and-pencil and be more dependent on language and thinking skills. As Gary Phillips suggests, there might even be an appropriate place for the venerable multiple-choice item.

The exact nature of the assessment task is determined by the goals stated for the curriculum and the assessment. Traditional assessment methods stress knowledge. Performance assessment stresses broader and more comprehensive responding. For example, after a biology unit on ecosystems, a student might be asked to set up a self-sustaining ecosystem in an aquarium. This task would obviously test more of the student's knowledge than would a multiple-choice item asking the child to identify which of five choices contains things necessary to sustain life.

The performance-based item would provide the student with a chance to show a more integrated understanding of the task. However, it is expensive in terms of staff and the child's time, and in other ways. Further, scoring of such a task is not as clear as scoring of traditional tests.

Another aspect is the cost. A true performance assessment of every child over every facet of the K-12 curriculum is beyond the total budget capability of Kentucky. This means that some sampling plan is necessary. Sampling will probably occur in terms of what is assessed from the curriculum, who is assessed on terms of students, schools and perhaps even school districts, and methods of assessments.

The law calls for test equating and establishing baselines of performance. Yet there are difficut technical problems that must be dealt with to accomplish this. The seeming simplicity of performance assessment can encourage avoidance of the thorny technical aspects. However, an assessment system must meet rigid criteria established for educational assessment systems if it is to be depended upon in the manner suggested by the the Kentucky Education Reform Act.

Schools will do continuous assessments of students performance and progress. These assessments may include a variety of methods of illustrating student progress. A very common one is the establishment of portfolios for each student. A portfolio is a folder that contains all the work the student has done (such as home work, traditional classroom tests, essays, and projects). Another, is to use group

problem solving approaches as well as traditional classroom tests and assignments to gauge performance. In fact, different methods might be appropriate depending on the goals stated for each sector of the curriculum, and goals for tying together various sectors, such as science and math, mathematics and language arts, etc. could be developed.

The continuous assessment that will be done in the schools will be important in the overall assessment of the school program. Further, teachers and administrators should have something to say about the design of this aspect of assessment.

Performance assessment will not be put into place all at once. The law suggests that this procedure will be done in stages. The first stage will be an interim testing program in 1991-92 in which a sample of students will be assessed to establish baseline data. A more complete implementation will follow in successive years. Regional assessment centers may be established to which samples of students might be brought for assessment activities. This implies that some aspect of the performance based assessment will be done outside of the regular school program and involve samples of students.

Other critical issues include consideration of individual students propensity to perform in different ways, regional differences in expression and experience, culture-fairness of certain kinds of tasks, and fairness to special populations. In addition, the match between the curriculum used in the classroom and the performance assessment method will require considerable attention. The amount of time school teaching staff have to prepare for and present assessment tasks, and their role interpreting and scoring should be explored as well.

Six goals have been established by the reform act for areas of student performance that will shape the school curriculum and the assessment. The goals are: (1) The application of basic communication and mathematics skills. (2) The application of core concepts and principles from all subject matter disciplines. (3) Becoming a self-sufficient individual. (4) Becoming a productive member of a family. (5) Thinking and problem solving. (6) Connecting and integrating knowledge.

The Kentucky Education Reform Act represents an integrated model of school reform. All of the major parts relate to each other in significant ways. For example, funding issues are tied to school performance. School performance is related to school-based decision-making, and so on. It appears that the performance-based assessment design will have implications for the kind and amount of information available to implement other parts of KERA. It is for this reason that it is critical that the performance assessment component is designed and executed well.

Conference Participants

Name	County	Name	County
Joyce Payne	Marion	Thomas A. Kelemer	Pineville Independent
Sue Gambrel	Bell	Joe Hignite	Perry
Pamela Saylor	Bell	Sheila Hullin	Woodford
Connie Wright	Bell	Max E. Gill	Woodford
Gene Tompkins	Bracken	Leon Smith	Washington
Don Turner	Danville Ind.	Daniel G. Bondurant	Washington
Una Farmer	Pulaski	Roy Chapman	Woodford
Colleen Lyons	Leslie	Ernest C. Taylor	Marion
JoAnn Fortney	Laurel	Steve Hutton	Beechwood, Ind.
John Chadwell	Middlesboro Ind.	Carl Hicks	Beechwood Ind.
Paul Spangler	Middlesboro Ind.	Brenda Anderson	Nicholas-Carlisle
Edwina C. Gibson	Monticello Ind.	Conrad E. Young	Clark
Delores Sapp	Montgomery	Joseph R. Jones	Clark
Barbara Amburgey	Montgomery	Marge Brackett	Clark
Sharon Noble	Hazard Ind. C	Jim Wells	Clark
Patti Price	Robertson	Ray Snowden	Estill
Doris M. Powell	Todd	Tom Hunt	Montgomery
Conley Manning	Williamsburg Ind.	Paul R. Spangler	Middlesboro Ind.
Jeanne Meacham	Robertson	Patsy R. Lester	Washington
Frank Howard	Fayette	Margaret Morrisey	Kenton
Mat Young	Lawrence	Donnie Collins	Hazard Ind.
John O'Brien	Ft. Thomas Ind.	Daniel L. Shull	Beechwood Ind.
Linda France	Jessamine	Richard P. Bowling	Leslie
Debbie Egan	Jessamine	Jennifer Hubbard	Kenton
Sandra Bailey	Knott	David S. Randolph	Boyle
Connie Aubrey	Jessamine	Jerry L. Leber	Boyle
James C. Hanks	Marion	Jack Tucker	Bourbon
David Hagan	Marion	Kirby M. Wright	Mason
Diane H. Evans	Marion	Jane H. Cline	Mason
Ruby M. Wright	Mason	Yvonne Slusher	Bell
Olen Taylor Collins	Leslie	Phyllis Whitaker	Magoffin
Anne Keene	Pikeville Ind.	Guy Strong	Clark
Mitchell Thompson	Marshall	William Bolton	Clark
Ken Rosenbaum	Jefferson	Pat Prestenberg	KY Valley Ed. Cooperative
Barbara Helm	Jessamine	Wanda C. Vice	KY Valley Ed. Cooperative
Ivan Cooper	Kenton	Jane Komara	Fayette
Sue G. Magee	Jessamine	Pat Nickell	Fayette
James W. Burks	Bourbon	Trulena Tate	Estill
James Whitaker	Jefferson	Carolyn Clark	Fayette
Barbara Stenzel	Jessamine	Janice Ledford	Ashland Ind.
Yvette Blair	Knott	Glenn A. Riedel	Ashland Ind.
Melissa Martin	Knott	Gayle Barnard	Elizabethtown Ind.
Wendy Lakes	Jessamine	Kenneth Barnard	Elizabethtown Ind.
Mike Norris	Jefferson	Brenda Priddy	Elizabethtown Ind.
Tom Mills	Boyle	Katherine Schloemer	Louisville Parochial Schs Jefferson
Michael H. Byers	Hardin	James L. Thomas	Estill
Ed Murphy	Fayette	Dale Brown	Carroll
Richard A. Steben	Fleming	Ed Watkins	Fayette
John Farler	Hazard	Ron Pelfrey	Fayette
Glen A. Hendrix	Leslie	Patricia Miller	Estill High
Karen S. Pravler	Mercer	Joseph Clark	Fayette
Johnnie Slone	Mercer	Bill Burch	Ashland Ind.
Paul K. Goodin	Pineville Ind.	Herb Conley	Ashland
Thomas Sizemore	Leslie	Lois Pritchett	Fayette
Paulette Offutt Kutter	Mercer	Betty Beardsley	Fayette
LaDonne Walker	Marshall	Jim Schlich	Fayette

BEST COPY AVAILABLE

Conference Participants

Name	County
Wilburn Nona	Leslie
Frank W. Baker	Leslie
Calvin Crawford	Leslie
Damon Huff	Leslie
Offina Shepherd	Leslie
Janie Ohr	Estill
Clayton Johnson	Estill
Patti Abney	Estill
Don F. Trivette	Fayette
Vikki Clemens	Campbell
Roger Meek	Boyle
Jay Carroll Miller	Hardin
Carolyn Murphy	Fayette
Ina Bradford	Carter
Jo Ashworth	Carter
Anthony Johnson	Augusta Ind.
Linda Roberts	Henry
Paula Martin	Henry
Tim Abrams	Henry
Jo Mitchell	Henry
Sharon Silvers	Henry
Nancy Sutton	Hazard Ind.
Jennifer Green	Jefferson
Carole Ramey	Jefferson
Martha Sandusky	Marion
R. B. Benningfield	Marion
Robert Davenport	Marion
Judy Gaddie	Marion
Jane Becker	Nicholas
Alfred M. Collins	Perry
Yvon Allen	Perry
Charles R. Campbell	Perry
Dennis P. Wooton	Perry
Bo McCreary	Model Lab School/EKU
Barbara Disney	Paris Middle
Patti Price	Robertson
Dorothy Lester	Russell Ind.
David M. Heath	Perry
Lois Hatfield	Pulaski
Michael Bromagen	Bath
Dulcie Hardin	Crittenden
Nancy Stalion	Crittenden
Susanne Hoagland	Madison
Donna Craft	Madison
Michael White	Pineville Ind.
Hazel S. Powell	Middlesboro Ind.
Sue W. Roberts	Marshall
Shelly Cohen	Jefferson
Teresa Petot	Paintsville Ind.
Bruce Bonar	Model Lab School
Jacqueline G. Vance	Model Lab School
Leanna Austin	Mayfield Ind.
Elsie Jones	Mayfield Ind.
Bonita Lykins	Mayfield
Carole Hancock	Pulaski
Reece Little	Jefferson

Name	County
Howard Hardin	Jefferson
Sandra Poe	Jefferson
Nan Franck	Jefferson
Sandra Lyvek	Scott
James Stewart McPeek	Pike
Roy Chapman	Woodford
James Clay	Pike
Cheryl Cox	Powell
Susan Miller	Powell
Martha Staker	Scott
Shelby Reynolds	Rockcastle
Twyla Hanna	Hardin
Sandy Greenwood	Hardin